

REASON, EMOTION, AND CONSEQUENCE:
MORAL PSYCHOLOGY AND KANTIAN IDEALS

Kiran Bhardwaj

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Philosophy.

Chapel Hill
2017

Approved by:

Thomas E. Hill, Jr.

Douglas MacLean

Ryan Preston-Roedder

Simon Blackburn

Bernard Boxill

© 2017
Kiran Bhardwaj
ALL RIGHTS RESERVED

ABSTRACT

Kiran Bhardwaj: Reason, Emotion, and Consequence: Moral Psychology and Kantian Ideals
(Under the direction of Thomas E. Hill, Jr.)

My project is a study of moral ideals, and particularly moral ideals as they play a role in Kantian ethics. It takes Kant's discussions of what he identified as ideals (the virtuous person, the ethical community, the highest good, and friendship) and uses those discussions to argue that Kantian ideals help Kantians respond to various features of our moral lives: as a concrete representation of the consequences of following the moral law (elicited through the imagination) that engages the emotions in a morally appropriate way.

The chapters of this dissertation provide: first, a general account of what ideals are and how they function, thanks to their employment of the imagination—explaining precisely how ideals orient and inspire us. Then within the domain of Kantian ethics, it reviews the Kantian debate about the supererogatory and provides a new account using Kantian ideals as a way to explore beyond the traditional explanation of actions that are good to do but not required. Chapters 3 through 5 each explore three of the Kantian ethical ideals: the ideal of self-perfection, the ideal of the kingdom of ends, and the ideal of friendship. These important themes in ethics—what it is to be a good individual, relationship, and community—are interestingly understood in the role of *ideals*, rather than solely in terms of duty.

ACKNOWLEDGMENTS

I was able to complete this dissertation, which was completed in fits and starts, thanks to my philosophical mentors. Tom Hill has been a generous, thoughtful advisor who has been an inspiration to me as a person and as a philosopher. Doug MacLean pushed me to be a better writer and a better thinker; I now hear his voice as I edit. I thank them both for reading more drafts of this dissertation than anyone ought to. The rest of my committee —Ryan Preston-Roedder, Simon Blackburn, and Bernie Boxill—gave me generous comments and steadfast encouragement. Roger Crisp was kind enough to read and thoughtfully comment on my dissertation—I continue to learn many things from him. I have also been helped by discussions with my peers: in particular, I am indebted to Jen Kling, whose advice was *always* on point, and Jordan MacKenzie, whose philosophical judgment is infallible, so far as I have evidence.

While I may have finished this project thanks to excellent philosophers, it wouldn't have been enjoyable without the partnership and love of Michael Boyle. My family and friends have also been incredibly important to me while finishing this dissertation. In particular, Mary and Vas Bhardwaj continue their perfect record of supporting my education; Meera Bhardwaj has been the ideal sister in every way; Raghu and Latha Ballal and Mariel Woods kept me grounded and loved (and fed!). To them and everyone else—thank you.

TABLE OF CONTENTS

INTRODUCTION.....	1
1. Criticisms of Kantian moral psychology.....	2
1.1 On Consequences.....	3
1.2 On Emotions.....	4
2. Kantian Appeals to Ideals.....	5
3. A Review of Kantian Ideals.....	7
4. Preview of Structure.....	8
Works Cited.....	11
CHAPTER 1: A CONCEPTION OF IDEALS.....	13
1. The Standard Account of Ideals.....	14
1.1 Coady's goal-like ideals.....	14
1.2 Brownlee's variation on Coady's account.....	15
1.3 Held v. Recognizable Ideals.....	17
2. How the Standard Account Fails.....	19
2.1. Stipulative v. Imaginative Ideals.....	19
2.2 Failures of the Standard Account.....	22
3. A Conception of Ideals.....	24
3.1 The Psychology of Imaginative Ideals.....	24
3.2 Our Relationship to Imaginable Ideals.....	31

4. The Imagination and its Pathological Uses.....	36
4.1 Kinds of Pathology.....	37
4.2 Guidance for Ideals.....	40
Conclusion.....	41
Works Cited.....	42
CHAPTER 2: THE SUPEREROGATORY, IMPERFECT DUTIES, AND IDEALS.....	45
1. The charge: Urmson's "Saints and Heroes".....	48
1.1 "Saints and Heroes".....	48
1.2 Urmson's criticisms of utilitarianism and Kantian ethics.....	50
2. The Kantian debate.....	52
2.1 Perfect and imperfect duties.....	53
2.2 Hill's account.....	55
2.3 Baron's criticisms.....	59
3. Kantian Supererogation and Ideals.....	65
3.1 Kant's conception of ideals.....	66
3.2 Our duties in the direction of ideals.....	74
3.3 Supererogation, imperfect duty, and ideals.....	76
Conclusion.....	81
Works Cited.....	83
CHAPTER 3: SELF-PERFECTION, SELF-IMPROVEMENT, AND THE KANTIAN VIRTUOUS AGENT.....	86
1. The Problem as it Arises in Kant's Moral Philosophy.....	87
2. Kant's Solution.....	91
3. Alternative Kantian Solutions.....	93

3.1 Self-Improvement View.....	94
3.2 Response to the Self-Improvement View.....	96
3.3 Faith in Humanity View.....	98
3.4 Response to the Faith in Humanity View.....	99
4. My View.....	100
4.1 The Lie of the Imagination.....	103
4.2 The Truth of the Imagination.....	108
Conclusion.....	110
Works Cited.....	112
CHAPTER 4: THE KINGDOM OF ENDS AS AN IDEAL.....	115
1. Kant on the Ethical Community.....	117
1.1 The kingdom of ends: as a community and as a legislative strategy.....	117
1.2 The kingdom of ends as an ideal.....	119
2. Interpretations of the Kingdom of Ends.....	123
2.1 A substantive ideal.....	124
2.2 A deliberative ideal.....	129
3. The Uses of an Ideal.....	131
4. The Imagination and Ideals.....	135
4.1 Mirroring and contagion.....	136
4.2 Testing the interpretations.....	138
Conclusion.....	144
Works Cited.....	145
CHAPTER 5: FRIENDSHIP AS A BRIDGE TO THE HIGHEST GOOD.....	147
1. Kant on Friendship.....	149

1.1 Kinds of Friendships.....	154
2. Moran on the Duty of Friendship.....	156
3. Moral Friendship and the Imperfect Duties.....	160
3.1 An Objection.....	164
4. The Ideal of Moral Friendship as a Bridge to the Highest Good.....	166
4.1 The Relationship between the Kingdom of Ends and the Highest Good.....	168
4.2 The Imaginative Ideal of Friendship and the Kingdom of Ends.....	169
4.3 Is the Ideal of Friendship Realizable?	171
4.4 An Objection.....	174
Conclusion.....	175
Works Cited.....	177

INTRODUCTION

My research project comes from a recurring theme in Kant's ethical writings: the ethical ideals regularly mentioned from the *Critique of Pure Reason* until the late ethical works. Other moral philosophers have meant something of great importance in their theory when they use the term; Kant's usage of the term is, comparatively, not taken to be central to Kantian ethics. Yet the *subjects* he identifies as 'ideals' are the heavy-hitters: the *virtuous person*, the *kingdom of ends* (or *ethical community*), the *highest good*, *friendship*. Each of these are important concepts in Kantian ethics. What significance might they bear, thought of as *ideals*?

In reviewing Kant's ethical writings and the work of other Kantians, I came to the conclusion that ideals can do a robust kind of work for us. In particular, they allow us to think in a very sophisticated way about moral psychology. Kantians agree on the importance of reason in the moral life. Yet there are other parts of our psychologies that provide us with incentives to act, including our concern for consequences and our emotions. Critics of Kantian ethics have suggested that Kant handles these features of our moral psychology badly (as I will discuss in the next section of this preface)—and think that he asks us to suppress the otherwise-potent influence of emotions and our concern for consequences when we are morally deliberating and acting.

Against these critics, I think we have a great deal of evidence to suggest that Kant acknowledged the significance of ideals in bridging to these other parts of our moral psychology. Ethical ideals *channel* those parts of our moral psychology—they get our emotions and our

representations of the consequences of our actions to fall in line with what reason requires.

Emotion and consequences (appropriately directed toward or engaged with Kantian ideals) are a meaningful part of human moral agency, even if they are secondary to the moral motivation of reason.

The particular ideals that Kant discusses (the virtuous person, the ethical community, friendship, and the highest good) are important aspects of a moral life. Many moral theories focus on duties and obligations of various sorts. However, ideals have a different kind of hold on us. An endorsable and inspirational account of the moral life would include these ways of imaginatively projecting to what such a person or world would be like. Kantian ideals, moreover, can do so in a way consistent with what the moral law requires. They do not ground moral action, but they point us toward what we can hope for in performing moral actions.

This Introduction will review some of the criticisms lodged against Kantian ethics and describe how Kantians have reacted to these concerns. Then, it will set up the major premise of the research project: that Kantian ethical ideals have a significant role to play in his moral psychology, and will set out these ethical ideals. Finally, it will preview the structure of the chapters of the dissertation.

1. Criticisms of Kantian moral psychology

Some critics argue that Kant's ethics fails to handle (in the right way) how we are oriented toward consequences and how emotion plays a role in our moral lives. I will take each of these lines of criticism in turn.

1.1 On Consequences

One obvious worry about Kant's ethics is that it fails to account for the importance of consequences, especially when the consequences radically change our ordinary intuitions about a moral situation. Consider the cottage industry on moral catastrophe in deontological ethics.¹ In particular, many critics have focused on Kant's comments about the 'murderer at the door'. These critics argue that Kant is so concerned with avoiding breaches of perfect duties, he ignores the fact that the direct consequences of being a truth-teller is that an innocent will be killed.

Kantians, of course, have responded to these concerns. Jens Timmermann, for example, reaffirms that consequences are not the foundational moral value of the theory (Timmerman 250). However, consequences do matter in a few ways in Kantian moral theory. First, the foreseeable consequences of the adoption of a certain maxim play a role in the thought experiment of the categorical imperative—the consequences are not decisive, but imagining the consequences will help us discover whether the proposed maxim is a moral one. (Timmermann 250-1). I might, for example, imagine the consequences if my maxim to tell a lie when it benefits me were to be universalized, using the Formulation of Universal Law. If everyone were to do so, the consequences would be that our practices of believing others would break down—and hence, the maxim is exposed as not moral. Second, the Categorical Imperative does not specify the means to a required or a suggested end. I will need to be thoughtful about the consequences of my actions to bring about my ends. For example, if I need to repay my debt to you, whether I am saving enough to pay you back matters (Timmermann 251). Moreover, with regard to my imperfect duties, I know I must promote the ends of my own perfection and the happiness of

¹ See, for instance, the discussion of moral catastrophe in *The Stanford Encyclopedia's* "Deontological Ethics" section. <http://plato.stanford.edu/entries/ethics-deontological/>

others. However, it is at my discretion for how I will pursue those ends. I might consider a wide range of consequences in order to decide about how to promote these ends (*MM* 4:446). Finally, the consequences of our actions are indirectly morally relevant because they affect our capacity to act morally (Timmerman 251). For example, we must secure our own happiness as an imperfect duty, because “when pressed by many cares and amidst unsatisfied needs” we would find it difficult to do our duty (*G* 4:399).

1.2 On Emotions

Another set of critics argue that Kantian ethics improperly dismisses our emotions.² For Kantian ethics, they argue, emotions are a *hindrance* to the possibility of morality.³ Bernard Williams’ arguments have been quite influential. In a series of essays⁴ he argued that Kantian morality devalues and estranges us from our emotions. In “Persons, Character, and Morality”, for example, he argued that the moral point of view charges us with impartiality. Yet this impartiality would objectionably require us to suppress our emotions towards those we care about when acting morally (Williams *ML* 2).

Feminist ethicists, for their part, have also argued that traditional moral theories (including Kantian ethics) improperly privilege the ‘rational’ over the ‘emotional’. Carol McMillan in *Women, Reason, and Nature* reads Kant as saying “that no action springing from

² Kate Moran admirably sets out these lines of criticism in *Community and Progress in Kant’s Moral Philosophy* (Moran 169).

³ For similar arguments, see Kekes, “Morality and Impartiality”; Stocker, “The Schizophrenia of Modern Ethical Theories”; Wolf, “Moral Saints”

⁴ “A Critique of Utilitarianism” and “Morality and the Emotions” in *Problems of the Self*; “Persons, Character, and Morality” in *Moral Luck*

natural inclination can have moral worth. For him, an action has genuine moral worth only when it is done solely out of duty, without any liking or preference for it” (McMillan 20-1).⁵

Of course, Kantians have given responses to these charges as well.⁶ Thomas E. Hill, Jr. argues that the impartiality thesis does not imply that “the emotional detachment of a judge is better than the compassion of a lover. No one is urged to live with his or her eyes fixed on abstract moral principles, still less with concentration on their justification from an impartial perspective” (Hill 133). Herman’s “Integrity and Impartiality” famously argued that Kantian moral theory does not entail the exclusion of the emotions for the sake of morality. Emotions still do not serve as the ground of morality, but Kantian ethics does not entail that “acting in the absence of or in opposition to emotions is necessarily desirable from a moral standpoint” (Herman “II” 233-50, 237; also see *MM* 4:456; Sedgwick 80). Duty must be ever-present as a limiting condition on what we do, but it does not need in every case be the primary motive (Herman “II” 237).

2. Kantian Appeals to Ideals

Emotions and consequences can be granted more potential in Kantian ethics. They are not foundational values for Kantian ethics, but they are an important part of human moral psychology—and Kant was sensitive to that fact. In particular, I think that Kant used the various

⁵ She cites the part of the Groundwork of the cold-hearted philanthropist to confirm her point. Also compare Baier, *Moral Prejudices*, on Gilligan, *In a Different Voice* (Gilligan 10, 65)?

⁶ Also see: Baron, “The Alleged Moral Repugnance of Acting from Duty”; Herman, “On the Value of Acting from the Motive of Duty”

ethical ideals—given his explanation of the role they play in our lives and deliberation—to confirm the appropriate place for emotions and our concern for consequences.

John Rawls (in his *Lectures on the History of Moral Philosophy*) set out a brief but promising lead that I will follow in this dissertation. Rawls acknowledges that it is hard to see how the moral law should connect with the final aims and purposes of human life, and with our moral psychology generally (Rawls 211). Rawls' answer is to look to Kant's discussions of ideals. Ideals can help to serve the requirements of the moral law, but also seem to capture possible consequences of doing so and arouse our emotions. Consider the kingdom of ends, for example. Rawls writes,

“Just as the representation, or an exemplar, of a morally worthy action done from a steadfast regard for the requirements of the moral law apart from all advantage to oneself uplifts the soul and arouses in us the wish that we could act in the same way, so, likewise, the ideal of a realm of ends (as the conception of a particular moral world) arouses in us the wish that we could be a member of that world.” (Rawls 213).

Rawls is not the only Kantian to indicate interest in the nature and function of Kantian ideals. Barbara Herman, in “A Cosmopolitan Kingdom of Ends” also looks towards this passage in Rawls. She notes that we don't know what the morally perfect person would be like exactly. Yet we do know that they would be “a person, mature and knowledgeable, in control of his desiring, who acts on certain principles” (Herman “CKE” 201). He would an *embodiment* of a regulative principle for our action (Herman “CKE” 201). As such, the ideal of the virtuous person can serve us incredibly well in our moral lives.

I intend to join these Kantian thinkers, and to continue to press the point that Kantian moral ideals play an important—and sometimes underappreciated—role in Kantian ethics. They can help Kantians respond to critics who think that Kantians do not appropriately acknowledge the force of emotions or consequences to creatures like ourselves. They can speak to us as the

complicated and internally inconsistent beings that we are, and help to buttress our activities as persons committed to the moral law.

3. A Review of Kantian Ideals

There are four ideals in Kant's ethical works: (i) the virtuous person, (ii) the kingdom of ends, (iii) friendship, and (iv) the highest good. Of these, my dissertation will focus directly on the first three. The highest good will be an indirect part of the discussions of the other three ideals.

We have already seen that there are textual reasons to think that Kant is sensitive to the power and significance of ideals. Consider the famous line in *Groundwork* about how the ideal of the kingdom of ends is intended to inspire and motivate us, to make morality more accessible "closer to intuition and feeling" (see Hill and Zweig 86; *G* 4:436-7).

If we extend our look to other ideals (like the virtuous person, the highest good) we see similar comments where Kant places weight on the excellence of ideals in our lives. Virtue, for example, "i.e. the firmly grounded disposition to fulfill one's duty strictly, is also *beneficent* in its consequences [...]. And if we consider the gracious consequences that virtue would spread throughout the world, should it gain entry everywhere, then the morally oriented reason (through the imagination) calls sensibility into play." (*R* 6:23n). Later Kant writes that if we imagine a person who sees his disposition constantly improve toward moral perfection's reflections on what he can hope of is "a glimpse into a *boundless* future which is, however, desirable and happy" (*R* 6:69). Likewise, the highest good is said to reconcile all the things that we desire: we desire virtue, and also happiness (*R* 6:5, 6:6n). It "can be interpreted as a symbolic representation

aimed merely at stimulating greater hope and courage and effort in achieving it” (*R* 6:133). We see, here, some reason to think that consequences and emotions are both in play when we reflect on the virtuous person or the highest good.

These ideals are constructed by reason. However, they also make the emotive parts of us, the consequence-oriented parts of us, fall in line with what reason requires of us. Ideals should not be taken as an *incentive* to be moral—Kant is more than clear that that is not their function (*R* 6:7). Yet these ideals are essential for directing our emotions, our desires, and other features of being empirical beings.

4. Preview of Structure

My project is to look for insight from Kant’s discussions of moral ideals (the virtuous person, the ethical community, the highest good, and friendship) and to argue that Kant’s framing of these concepts *as* ideals is an underutilized resource for Kantians. I argue that Kantian ideals help Kantians respond to those who think that Kantian ethics is inadequate because it fails to treat important features of our moral psychology (especially consequence and emotion) with due consideration.

The first task, in order to do this, is to construct an account of ideals. **Chapter 1** takes this task on. The Standard Account of ideals in the broader ethics literature articulates characteristics of the ideals that play a central role in human life. However, the Standard Account is unable to explain why ideals are the kinds of things that we like to imaginatively engage with, nor explain why ideals function (and malfunction) in the way that they do. This chapter presents a conception of ideals as imaginative representations of excellence for an object or characteristic

of an object. The account takes characteristics of the imagination (‘immersion’, ‘contagion’, ‘mirroring’) to help us understand how ideals can orient us, and why they can be dangerous or delusional.

With a conception of ideals in hand, we can follow through to the next chapters. Urmson rightly emphasized that in moral life and good moral theory, we distinguish between what one *has* to do and what is commendable, but not required. **Chapter 2** engages with this topic. The traditional Kantian explanation of the category of actions that are good to do, but not required (as seen in the work of Thomas E. Hill, Jr.) relies on the nature of imperfect duties. While we are always required to perform perfect duties, we have some latitude in how we fulfill our two imperfect duties (that we perfect ourselves and that we promote the happiness of others). Not only may we fulfill these imperfect duties in a variety of ways, we also may sometimes refrain from acting in ways that fulfill them, so long as we regularly promote those two ends. When a person promotes one of these ends when they could permissibly have refrained, and the action is admirable (that is, their motive was good), it can be considered “supererogatory” on a Kantian account. I argue that this supererogation-via-imperfect duty account is not sufficient, and give a new account of Kantian supererogation. This account relies on the significance of Kantian ethical ideals.

Chapters 3 through 5 each inspect three of the Kantian ethical ideals. **Chapter 3** explores the ideal of self-perfection—that we should, in some sense, ‘be all that you can be’. In moral theory, there are three claims that are each convincing but are mutually inconsistent: (1) that you ought to “be all that you can be”, (2) that it is not possible for you to be all that you can be, and (3) that ought implies can.⁷ Kant, surprisingly, rejected the plausible claim (2). Some Kantian

⁷ This is a quick rendering of the argument, and will be more carefully set up in Chapter 3.

thinkers have thought that we should instead reject claim (1). I argue that neither the rejection of (1) or (2) is satisfying, and we have good reason not to reject (3). Instead, I show how the moral imagination holds the key to the apparent puzzle raised by our duty of self-perfection.

Chapter 4, in turn, looks at the kingdom of ends as an ideal. The chapter has two parts. In the first, I compare two readings of the kingdom of ends. On the one hand, the kingdom of ends may be what we would call a “substantive ideal”: we imagine a perfect moral community, and we try to make our real-world community more and more like the perfect one. On the other hand, some interpreters have thought that the kingdom of ends is better-read as a “deliberative ideal”: we can imagine that we are acting as the ideal *deliberator* (taking on the point of view of a member of that community who is responsive to various features of the world). The second part of the chapter tests how the two interpretations of the kingdom of ideals can work as imaginative ideals, and shows how we can understand each interpretation more strategically.

Chapter 5 focuses on the ideal of friendship. This chapter examines Kate Moran’s interpretation of the duty of friendship, and while largely in favor of her argument, suggests that she has missed two important parts of what the ideal of friendship does. (1) Friendship helps us practice our duty of moral perfection (as Moran indicates) but it also helps us practice our duty of beneficence and (2) that it is an ideal that Kant says is achievable means that it can serve as a bridge to the highest good—a good that is characteristically unachievable.

WORKS CITED

- Alexander, Larry and Moore, Michael, "Deontological Ethics", *The Stanford Encyclopedia of Philosophy* (Spring 2015 Edition), Edward N. Zalta (ed.), Online: <http://plato.stanford.edu/archives/spr2015/entries/ethics-deontological/>
- Baier, Annette. *Moral Prejudices*. Cambridge, MA: Harvard University Press, 1994.
- Baron, Marcia. "The Alleged Moral Repugnance of Acting from Duty" in *The Journal of Philosophy*, Vol. 81, No. 4 (April 1984), pp. 197-220.
- Gilligan, Carol. *In a different voice: psychological theory and women's development*. Cambridge, MA: Harvard University Press, 1993.
- Herman, Barbara. "A Cosmopolitan Kingdom of Ends" in *Reclaiming the History of Ethics: Essays for John Rawls*, edited by Andrews Reath, Barbara Herman, and Christine M. Korsgaard. Cambridge: Cambridge University Press, 1997.
- Herman, Barbara. "On the Value of Acting from the Motive of Duty" in *Philosophical Review*, Vol. 90, No. 3 (July 1981), pp. 259-382.
- Herman, Barbara. "Integrity and Impartiality" in *The Monist*, Vol. 66, No. 2 (April 1983), pp. 233-50.
- Hill, Jr., Thomas E. "The Importance of Autonomy" in *Women and Moral Theory*, edited by Eva Feder Kittay and Diana T. Meyers. Totowa, NJ: Rowman & Littlefield, 1987.
- Hill, Jr., Thomas E. and Arnulf Zweig, "Editors' Introduction" to the *Groundwork for the Metaphysics of Morals* [Gr], translated by Arnulf Zweig, edited by Thomas E. Hill Jr. and Arnulf Zweig. New York: Oxford University Press, 2002.
- McMillan, Carol. *Women, Reason, and Nature*. Oxford: Blackwell Publishing, 1982.
- Moran, Kate. *Community and Progress in Kant's Moral Philosophy*. Washington, DC: Catholic University of America Press, 2012. ProQuest ebrary. Web. Accessed January 28, 2017.
- Kant, Immanuel. *Groundwork for the Metaphysics of Morals* [G], translated by Arnulf Zweig, edited by Thomas E. Hill Jr. and Arnulf Zweig. New York: Oxford University Press, 2002.
- Kant, Immanuel. *Metaphysics of Morals* [MM], translated and edited by Mary Gregor. New York: Cambridge University Press, 1996.
- Kant, Immanuel. *Religion within the Boundaries of Mere Reason* [R], edited by Allen Wood and George di Giovanni. New York: Cambridge University Press, 1998.

- Kant, Immanuel. "On a supposed right to lie because of philanthropic concerns" in *Grounding for the Metaphysics of Morals; with On a supposed right to lie because of philanthropic concerns*. Indianapolis: Hackett, 1993.
- Kekes, John. "Morality and Impartiality" in *American Philosophical Quarterly*, Vol. 18, No. 4 (October 1981), pp. 295-303.
- Rawls, John. *Lectures on the History of Moral Philosophy*, edited by Barbara Herman. Cambridge, MA: Harvard University Press, 2000.
- Sedgwick, Sally. "Can Kant's Ethics Survive the Feminist Critique?" in *Feminist Interpretations of Immanuel Kant*, edited by Robin May Schott. University Park: Pennsylvania State University Press, 1997.
- Stocker, Michael. "The Schizophrenia of Modern Ethical Theories" in *The Journal of Philosophy*, Vol. 73, No. 14 (August 1976), pp. 453-66.
- Timmermann, Jens. "Kantian ethics and utilitarianism" in *The Cambridge Companion to Utilitarianism*, edited by Ben Eggleston and Dale E. Miller. New York: Cambridge University Press, 2014.
- Williams, Bernard. *Problems of the Self*. New York: Cambridge University Press, 1973.
- Williams, Bernard. *Moral Luck*. New York: Cambridge University Press, 1981.
- Wolf, Susan. "Moral Saints" in *The Variety of Values: Essays on Morality, Meaning, and Love*. Oxford Scholarship Online (Online Publication: December 2014). New York: Oxford University Press, 2015. Accessed January 28, 2017.

CHAPTER 1: A CONCEPTION OF IDEALS

Ideals are a pervasive part of commonsense morality and feature centrally in many normative ethical theories. They also are embedded in our ordinary lives: they are a rallying point for political and social rhetoric, help to coordinate groups (e.g., journalistic or scientific ideals), and inspire and direct individuals making choices about their lives (e.g., the WWJD bracelets from the 90s). It may then come as a surprise that there is scant discussion in contemporary moral philosophy of the characteristics of ideals, and very little on the reasons why ideals are so effective for us, and how they can become dangerous or delusional. This lack of attention is very unlike philosophers' close work on other important normative concepts. In this dissertation, I will contend that a close study of ideals can help to illuminate many features of our moral practices, especially within Kantian normative ethics. In this chapter in particular, I will present the Standard Account of ideals, and use contemporary psychological and philosophical theories of the imagination to offer a better account of the nature and function of ideals.

This chapter will first review the literature on ideals. In currently available definitions of ideals⁸, they are goal-like objects which differ from goals in various important ways—by being highly valuable, the object of pursuit over a full life, unrealizable, and so on. Then it will turn to a discussion of the distinction between *stipulative* and *imaginative* ideals. This distinction allows us to recognize the philosophically interesting features of imaginative ideals. An imaginative

⁸ In particular, the work of C. A. J. Coady and Kimberley Brownlee.

ideal is an *imaginative* representation of excellence for an object or characteristic of an object. This application of the imagination to the objects we personally value will be what results in the conventional characteristics of ideals. The remainder of the paper will serve as a discussion of the complex relations between the proper use of the imagination and well-functioning ideals.

1. The Standard Account of Ideals

The literature on ideals is small and will not necessarily be familiar to all readers. To show how the account of ideals I will offer is distinctive, I will first set out the accounts offered by C. A. J. Coady and Kimberley Brownlee, which I will call the “Standard Account”.⁹ Both accounts focus on the ideals that have a distinctive kind of importance in our lives.

1.1 Coady’s goal-like ideals

Coady’s (2008) *Messy Morality* is an expanded and modified version of his 2005 Uehiro Lectures, the third chapter of which, he says, “seeks to restore the concept of ideals to an important place in philosophical discussion, and to give it a particular pertinence in the discussion of politics” (Coady viii).

Coady gives one of the more robust analyses of ideals available in the contemporary literature. He notes how ideals are similar to ‘goals’, but suggests that ideals differ from more mundane goals in at least four ways. *First*, they are more comprehensive and general than ordinary goals. *Second*, they are highly estimable. *Third*, they are ‘pervasive’ and ‘constitutive’

⁹ For other discussions of ideals, see Rescher, *Ethical Idealism*; Emmet, *The role of the unrealizable*. Strawson, “Social Morality and Individual Ideal”.

(terms that he takes to mean that (i) ideals impact many parts of our lives, and (ii) by seeking to live the ideal, the ideal comes to exist to some degree (Coady 52, 57). *Fourth*, ideals are unrealizable (Coady 51-2). Unrealizability could be cashed out in one of two ways: they could be unrealizable because of the limits of an individual's capacity or circumstances, *or* they could be absolutely unrealizable (Coady 55). (That is, Coady picks out a distinction between a personally impossible ideal and an empirically impossible ideal). Taken together, these features of ideals provide a dynamism to moral life: providing "vitality and direction" to what we do (Coady 53).

How can ideals guide us? Coady thinks ideals, being characteristically unrealizable, are often not to be aimed at. Instead, they (i) inform us and provide direction for our behavior, (ii) are a more functionally efficient referent than a more complex set of instructions, and (iii) perhaps can help us rightly alter our beliefs about what is possible (Coady 61).

1.2 Brownlee's variation on Coady's account

Brownlee's (2010) paper "Moral Aspirations and Ideals" is an attempt to vindicate not only ideals but also aspirations, which she defines as an attitude of commitment to, striving for, or desire for an ideal (Brownlee 241, 243). Her account of ideals is modeled on Coady's. Ideals, on her view, are "models of excellence or conceptions of perfection around which we can orient our thoughts and conduct" (Brownlee 243).

Brownlee expands Coady's account from four to five categories and differentiates her account from Coady's on several points. She agrees that the kind of ideals under discussion are comprehensive—and the way to tell that a given ideal is comprehensive is that it could plausibly be the core focus of a meaningful life (Brownlee 243). She also agrees that ideals are admirable: they represent things that rank very highly as goods. Brownlee takes an objective view of

admirability. She acknowledges that some so-called ‘ideals’ have been thought to be genuine ideals by some adherents (e.g., the Nazi ideal of racial purity), thanks to how they were wrongly perceived to be admirable (Brownlee 243). Genuine ideals, in contrast, are worth admiration, as is a person’s success in realizing aspects of the ideal (Brownlee 244).

Brownlee separates out, into three parts, Coady’s categories of ‘pervasiveness’ and ‘constitutiveness’ (which Coady had treated together). Brownlee argues that an ideal has a pervasive effect on our lives in two different ways, and that constitutiveness is a yet another feature of ideals. By saying that ideals are pervasive, she means, *first*, that a given ideal can be realized in a multiplicity of ways, and has a multiplicity of constitutive elements. For example, there are many ways to be a musical virtuoso, and different physiques best suit kinds of athletic excellence (Brownlee 244). *Second*, she means that a range of domains of our reasoning are shaped and influenced by our chosen ideals (Brownlee 244). Next, ideals are constitutive. Here she again agrees with Coady: the core behavior undertaken to cultivate an ideal is constitutive of that ideal itself, and is not merely an instrumental means for pursuing it (Brownlee 245).

She agrees with Coady’s account that ideals are unrealizable. Brownlee, like Coady, thinks that ideals must be ‘beyond us’ in some way. Unrealizability, however, comes in degrees, from personally unrealizable (being ‘like Mike’ is an unrealizable ideal to many; but realizable for Michael Jordan) to unrealizable for all persons at any circumstances, or for all persons at all times (Brownlee 245-6). However, ideals are not unrealizable to the point at which they defy the imagination (Brownlee 246). She writes that “ideals, as conceived of here, originate in the use of the imagination. They not only arise from reflection upon how best to push beyond our apparent limits, but also, consequently, their cultivation allows us ‘to contemplate value possibilities that transcend the restrictive confines of the real’ [Rescher 83]” (Brownlee 246).

Brownlee's final characteristic of ideals is that they are the proper objects of aspirations. Aspirations are attitudes of commitment to, striving for, or deep desire for an ideal as a model of excellence (Brownlee 243, 247). It is important to have meaningful moral aspirations to do and be better than we have been. A life without such aspirations, Brownlee says, would not be a meaningful one for moral agents (Brownlee 249-50).

I take the Coady-Brownlee account to articulate the Standard Account of ideals, where the focus is on those ideals that are of central concern in human lives. Ideals, on this view, are *admirable, pervasive, constitutive, unrealizable, and the objects of aspiration*.

1.3 *Held v. Recognizable Ideals*

It will prove to be helpful to make one small extension to how Brownlee and Coady articulate the Standard Account. A distinction that is only implicit in Coady's and Brownlee's work is an acknowledgement of the ways in which ideals are often parochial.¹⁰ Ideals, like opinions and preferences, can be *held* or they can be *recognized*. Since we exhibit different, and more engaged, kinds of behaviors towards our held—or 'personal'—ideals, it is worth identifying the differences in more detail than is available in the Brownlee and Coady accounts.

Let us first approach this distinction by example. I happen to hold the ideal of an oboist—I play the oboe, I want to become better, and I find reflecting on what the best kind of oboist would be helps to play a role in my practical reasoning and to shape my deliberations. Those who know me well know this fact about my ideals. They might to a greater or lesser extent know

¹⁰ Brownlee uses the phrase "chosen or acknowledged ideals" to capture this distinction (Brownlee 243; also see Brownlee 244). I do not follow Brownlee's language because not all of the ideals we happen to subscribe to are ones that we have *chosen*. Coady speaks of 'insiders' (those who hold an ideal) and 'outsiders' (those who do not) to capture this same distinction (Coady 68). Rescher discusses those ideals that are mine, and those that I would welcome others to share with me (even if I cannot expect or demand it) (Rescher 122-3).

what is entailed by the ideal. They *recognize* that ‘the ideal oboist’ is one that I or others care about. Nonetheless, the ideal does not similarly factor in their practical reasoning or help to shape their deliberations.

Thus, a *held ideal* represents what a particular person believes to be the highly-valuable features of a certain object or of one of its characteristics within a particular, identifiable context. The personal ideals which we hold *must* be concerned with ourselves or something that we share in. For example, I could hold an ideal of what I or what my community could be like (‘my ideal self’, ‘my ideal society’), an ideal for a social role I play (‘teacher’, ‘philosopher’, etc.), or an ideal human characteristic that I only imperfectly share in (‘patient’, ‘graceful’). One might hold an ideal on someone else’s behalf, but this only can be a personal ideal if the holder is engaged with that other person’s life. For example, I am concerned with my mother’s ideal of being a good parent—not only because I am her child (and in many ways the result of her parenting), but also because I anticipate that I will take on, either consciously or unconsciously, many of her habits in my own parenting.

Contrastingly, a *recognizable ideal* is where one recognizes something as an ideal that someone could hold *or* as an ideal of a kind. The recognizer might be familiar with those who claim it as an ideal, or can easily imagine what it would be like to hold it as an ideal. For example, she might know that someone from a different religion holds an ideal that is worthy of respect or admirable in some way—even though it is not an ideal she shares in. Since my talents in the visual arts peaked at stick-figures, the ideals held by artists are the kinds of ideals that I recognize as ideals. I can imagine holding them, but do not hold them myself.¹¹

¹¹ As a matter of how we speak, we do refer to ‘ideals’ that *no one* may hold, but which capture the excellences of an object of a certain kind. For example, we might be able to identify the features of “the ideal parasite”, even though it is unlikely that anyone holds this as an ideal—and certainly is not one of the special and significant ideals that Coady and Brownlee are concerned with. There are other ideals that some people *do* hold, though they are vicious or improper

2. How the Standard Account Fails

Ideals are a multivocal concept. As Coady writes, “The term itself eludes tight definition since it provides a mode of description that operates on a range of other moral categories in such a way that something will be an ideal in one context and not in another” (52). This means that for each of the features of ideals listed in the Standard Account, we could come up with a counterexample: something we think of as an ideal, but is not pervasive, or is realizable (or so on). Yet Coady’s and Brownlee’s purposes were to capture the main features of ideals that play a distinct and important role in our lives, and their accounts are successful in that regard.

However, the Standard Account fails to give a fully adequate account of ideals. In order to understand why, this section will *first* introduce a distinction between two varieties of ideal, stipulative ideals and imaginative ideals, and will *second* show that a more successful account of ideals must speak to the features of imaginative ideals. In Section 3, I will give an account of ideals that makes these additions.

2.1 Stipulative v. Imaginative Ideals

If we look at how we deploy the term ‘ideal’, it seems like there are two varieties of ideal: stipulative ideals and imaginative ideals. I will describe both, and suggest that the latter is especially philosophically interesting. A *stipulative ideal*¹² is one in which the representation includes a stipulation of the highly-valuable features of that object or characteristic of an object

ideals. I have some conception of what it would take to be an ideal counterfeiter or bank robber, even what it would be to *hold* such an ideal. However, on the Standard Account—since what matters is *genuine admirability*—we can dismiss these as not a matter under concern.

¹² I do not intend, by use of the word ‘stipulative’, to suggest that stipulative ideals are in some sense arbitrary or at will. The sense of the term is intended to pick out the ways in which these are ideals formed using propositions or concepts, and can meaningfully be put into words. In contrast, imaginative ideals are in some ways beyond description because they capture (in some sense) ‘what it would be like’ for those excellences to exist in an object or person.

that we are concerned with. That is, it is a conceptual assessment of features of excellence that hold for the target. Recognizable ideals are quite often, though not necessarily, stipulative; and our held ideals could be as well. In such a case, I might understand what the ideal entails, but I do not imagine what it would be like if the ideal were to come about. I will interchangeably refer to stipulative ideals as ‘standards of excellence’. For example, a stipulative ideal of a legal system might list features such as that disputing parties have appropriate and equal standing in the legal process, that it is accessible to all, that it adequately acts to uphold the laws, that the decision is made by an impartial judge or judges, that it is not susceptible to corruption, and that the results are fair and acted upon accordingly. We could articulate a checklist, of sorts, for the stipulative ideal of a legal system.

In contrast, an *imaginative ideal* is an imaginative representation of the highly-valuable features of that object or characteristic we are concerned with. This could be a full or partial imagining of that object or characteristic. So the imaginative counterpart to the stipulative ideal of a legal system might be one in which we *imagine* persons in a room, presenting sides of cases, being respectful of others, and where decisions are made carefully and thoughtfully. Or perhaps its imaginative counterpart will be one a little closer to reality: where we imagine flawed and imperfect people engaging with a legal process that does not always produce a just outcome. Nonetheless, it still involves imagining people making a genuine effort to seek justice.

What distinguishes standards of excellence from imaginative ideals is that an imaginative ideal must import some feature of ‘what it is like’, as brought about by the imagination. It cannot be a mere conceptualization of excellence for the object. For physical objects, it might include perception-like imaginings (e.g., ‘the ideal apple is slightly tart and crunchy’). Or it might involve imagining the characteristic beliefs and desires and mental life of another kind of person

(e.g., trying to get in the mindset of the ‘ideal musician’), or one’s own beliefs and desires and mental life if the ideal were to come to pass in my life (e.g., my life would be like *so* if I got my ideal job).

We can flip back and forth between these two ways of engaging with our ideals. At one point, I might regard one of my held ideals as a list of characteristics; at another, I might imagine what the ideal would be like if it were a reality.¹³ Both kinds of ideals in our repertoire are useful and motivating to us. Nonetheless, there is something special and important about the function of imaginative ideals.

Imaginative ideals motivate us in a different, and (for many of us) even more effective manner than their stipulative counterparts. An explanation for this difference because of how the imagination functions to mimic beliefs, desires, and perceptions (this topic will be discussed in more detail in Section 3). The imagination hooks up to our existing beliefs, desires, and responses to perceptions. Imaginings elicit visceral reactions in much the same way that perception, belief, and desire can. An imaginative ideal, then, comes with a concomitant affective and physical response.¹⁴ Those features can explain the special kind of motivation that our held ideals have on our emotions and physical responses.

The stipulative ideal may be able to—in a dry and precise way—articulate all the aspects of excellence of an object or characteristic of an object. Stipulative ideals can be intellectually motivating, in the same way that rules can motivate us to act. However, the imaginative ideal can

¹³ This might be similar to Alvin Goldman’s distinction between ‘supposition-imagination’ from ‘enactment-imagination’: the former is ‘conceptual’ and is formulated with a “that”-clause, and the latter involves creating a facsimile of a mental state in one’s mind (Goldman 41-2), or Currie and Ravenscroft’s distinction between ‘propositional imaginings’ and ‘perceptual imaginings’ (Currie and Ravenscroft 12).

¹⁴ The somatic responses to imaginings have been explored by Damasio et al. (1991).

motivate us by mimicking what it would be like for those qualities of excellence to be in the world. So we will want an account of ideals to be able to not only recognize this distinction, but be sensitive to the ways that the imagination provides a special function for ideals.

2.2 Failures of the Standard Account

I propose that the Standard Account, as it stands, is incomplete in a certain way because it does not explain the ways in which imaginative ideals function. Now, the claim that we must recognize the importance of the imagination in order to understand ideals is not novel. Both John Dewey and George Santayana thought that ideals are closely related to the imagination.¹⁵ Brownlee, too, thinks that imagination is an important part of understanding why ideals are a meaningful part of our lives.

Brownlee writes that ideals both originate in and are objects of the imagination (Brownlee 246, 257)—which is why, even if they are unrealizable, ideals stay within the boundaries of the imaginable (Brownlee 241). As such, they can assist us to understand both how we can move beyond our apparent limits (Brownlee 246) and expand our sense of what even might be possible (Brownlee 241). Aspirations (the attitudes that we have towards ideals) are, correspondingly, imaginative acts (Brownlee 246). Ideals “[take] the imagination by storm” and in so doing, result in aspirations (a ‘deep commitment’ and ‘longing’) (Brownlee 253).

I agree with Brownlee in the ways that she indicates that ideals are steeped in the imagination. Yet the function of the imagination with regard to these ideals is still rather

¹⁵ Santayana writes, “all observation is observation of brute fact, all discipline is mere repression, until these facts digested and this discipline embodied in humane impulses become the starting point for a creative movement of the imagination, the firm basis for ideal constructions in society, religion, and art.” (Santayana vii-ix). Dewey also thought it is “generally” recognized that the imagination is connected intimately with ideals (Dewey 19). In particular, he writes that “The connection between imagination and the harmonizing of the self is closer than is usually thought. The idea of a whole, whether of the whole personal being or of the world, is an imaginative, not a literal, idea.” (Dewey 18).

mysterious on her description. What we need is a conception of ideals that is rooted in contemporary work on the imagination. In particular, it should be able to do the following three things.

First, the account ought to explain our tendency towards imaginatively engaging with our ideals. The fact that we can engage with ideals both stipulatively *and* imaginatively means that holding an ideal does not necessarily exercise the imagination. However, as Brownlee points out, we like to engage with them imaginatively and we think there is something special about doing so. So what is it about *imaginative* ideals that we find compelling?

Second, the account ought to be psychologically sophisticated. While Brownlee nods towards the ways in which the imagination gives ideals force, there is a decided lack of detail in Brownlee's account. *Why* is it that imaginative ideals have the ability to function in the ways that they do? Section 3 will take on these first two questions.

Finally, our account must be able to explain the pathologies of ideals. Many philosophers and ordinary users have worried about ideals' ability to be dangerous or delusional. Coady thought that ideals had this tendency only in the case when we pay single-minded attention to them (Coady 63). Yet when we are 'caught up' in problematic idealistic thinking, we often find it hard to extricate ourselves—so I suspect it is not merely a matter of attentiveness. If my account can explain the phenomena of pathological uses of ideals, then it will be stronger than the Standard Account. I will take this question on in Section 4.

3. A Conception of Ideals

In this section, we have two tasks ahead of us. First, we must ask what is it about *imaginative* ideals that we find so compelling. Second, we must ask why ideals function in the way that they do. To assist us in these tasks, I suggest that we borrow some insights from the philosophical and psychological literature on the imagination, and connect those insights to the features that we take ideals to characteristically exhibit.

3.1 *The Psychology of Imaginative Ideals*

What does the imagination do? Generally, it enables “us to project ourselves into another situation and to see, or think about, the world from another perspective.” (Currie and Ravenscroft 1). The imagination is also, importantly, directed to what it would be for x to be the case, whether or not it is in fact the case (Schellenberg 498). As such imaginings are not *quite* perceptions or beliefs or desires, but are relevantly like those states. They can mimic each of those states and sometimes substitute for them (Currie and Ravenscroft 11).

Some imaginings are belief-like: I might imagine that the floor is lava; perhaps which would then be accompanied by real feelings of the fear of being burnt. Some imaginings are desire-like: I might imagine what emotions I would feel if I had an archenemy—hatred, contempt, and so on. And other very common imaginings are perception-like. I could imagine what it is like to see an orange, long-haired cat sitting on the table in front of me: I bring to mind an analogue of perceptions of cats that I have seen sitting, and have an imaginative experience that mimics those past experiences.

When reading fiction, or engaging in games of make-believe, we become immersed in an imaginative world. As Susanna Schellenberg points out, “The most relevant characteristic of

imaginative immersion is that the subject does not consciously think about the fact that she is imagining” (Schellenberg 507). The more we immerse in an imagining, the more we invest—and so immersion can be beneficial to learning, allowing us to see perspectives that we might not have otherwise and to practice new strategies (Schellenberg 508).

We imaginatively immerse in our personal ideals. We *imagine*, for example, what Jesus would do when we’re prompted by our WWJD bracelet. We imagine Jesus, with all his exemplary characteristics, in our circumstances—acting in a way that is presumably much better than the way we are tempted to act. (Or we try to step inside the head of such a morally exemplary person, and aim to think of what they think about). We imaginatively engage with Mandela’s “ideal of a democratic and free society”¹⁶— we begin from the people (including, perhaps, ourselves) who we know are denied equal opportunities and imagine what our world would be like if it were more like the ideal.

These imaginative ideals are relevantly like ordinary experience; perhaps we would have perception-like imaginings of seeing Jesus act, or our desires would be sparked by immersing in what it would be like to live in a democratic and free society. The engaging quality of imaginatively immersing in an ideal that speaks to our dearly-held values is why we like immersing in feel-good movies or happy endings.

Yet there is more to the psychology of imaginative ideals. Two other features of the imagination—mirroring and contagion—are also relevant to understanding how ideals function, and why.

¹⁶ Mandela, Nelson. “An ideal for which I am prepared to die”— Nelson’s statement at the opening of his trial on charges of sabotage in 1964.

3.1.1 Mirroring

Not only does an imagining mimic beliefs, desires, or perceptions, I can also get feedback from imaginings similar to what I would receive from an actual belief, desire, or perception. We imagine in ways that “mirror” reality; that is, when we pretend X is Y, X is taken to have the effects and features *within* the pretense that Y is in reality believed to have (Gendler 142-3). I might be forced to reckon with the natural implications of taking a course of action; when I imagine that I’m pouring tea and have missed the cup, I will ‘mirror’ reality and imagine that the table is now wet. So from imagining a certain state of affairs or object, I might be forced to attend to both realistic physical and psychological consequences.

We can have imaginings that, if represented vividly, can garner similar physical and emotional reactions to actual perceptions. These kinds of responses can be easily reconstructed: imagine a ball coming quickly toward your face, and you may flinch. Vividly think about the feeling and smell of stepping in a pile of dog poo; we feel disgust. This is partly because of the way that our motor system resonates. Consider, for example, how world-class athletes use imagining to boost their athletic performance, thereby improving their motor and motivational skills during a stressful situation. The motor pathways used in the motor performance result in more rapid response time and more accurate performance if one images what it feels like to perform well (Decety and Stevens 7).

Imagination can also be a tool to become more transparent to ourselves. Notably, I may recognize corresponding beliefs or desires or emotions that I might not have otherwise recognized that I have. A person might realize how vulnerable she feels, for example, by imagining herself in the position of a fictional character—and recognizing features of that character’s world in their own life.

Since the responses we have to our imaginings are *ours*, we ought to take our tendencies in imagining as more often indications about ourselves rather than insight into other minds (after all, we populate our imaginings through our own past experiences and psychological makeup). Nonetheless, we use imaginings of another person's mind, as well as projection from our own selves, and other strategies, to be able to gain insight into other minds (Myers and Hodges 281). We do best when we are motivated and skilled enough to create a coherent mental representation of the person we are targeting (Myers and Hodges 285). Understanding others' ideals might also lend us more insight into the kinds of responses they feel as well.

The mirroring feature of the imagination can also help us in our planning and decision-making. If we take time to imagine what we should do in a planning context, it gives us an explorative mind-set, which allows us to identify more means to whatever end we have set (Faude-Koivisto et. al. 80). Mentally simulating various courses of action might allow a person to get clear on their options moving toward a goal.¹⁷

Mirroring can help us explain a special characteristic of our engagement with ideals: the way in which ideals 'harmonize' various conflicting features of the object in the world. We know that there are conflicts in desirable features of an object, because to satisfy some to a high degree would rule out others. For example, the 'ideal community' is a way to think about how a community like ours, which is full of people with mutually inconsistent plans and projects and whose actions conflict with one another's, might be able to coordinate well. Each person, as well, has plans and motivations that conflict with the others. If I think of my ideal 'best self',

¹⁷ However, the researchers suggest that once we have selected a particular option, we benefit more from forming 'implementation intentions' than continuing to imagine. Implementation intentions are "if-then" plans which alert me to situational cues, and activation of a planned behavioral response (Faude-Koivisto et. al. 82)). While the exploratory mindset of the imagination is more open, allowing us to explore our options, we also can be distracted by those options from acting on our plans at this stage of execution.

then I might be able to help reconcile between these inconsistent parts of myself. *Mirroring* explains why ideals can perform this characteristic function—while ideals break away from what is actually true, they accurately reflect features of the object. Because imaginings mirror, they can allow us to test and adjust those features and to put them into balance. At its best, harmonization can mirror realistic physical and psychological consequences.

3.1.2 Contagion

We often take imaginings to be “quarantined” (Tamar Gendler’s term) or “encapsulated”. An imagining is quarantined when we recognize that the pretense episode *is* a pretense episode, and anything that we simulate is a mere simulation (Gendler 7). When we are vividly immersed in an imaginative episode, emotions and beliefs that are responses to what we are imagining might ‘escape’ quarantine. (Consider a person who is angry at a fictional character, and her anger colors how she treats those around her even after she has put the book down.)

This phenomenon is called “contagion”. We recognize a given imagining *as* not reality-reflective, yet we still can’t help but respond to it as if its content *were* reality-reflective (Gendler 242). This is different than responding within the context of the pretense. For example, a psychology study showed that test subjects were reluctant to eat from a jar they had themselves filled with sugar and then labeled as “cyanide” (Gendler 246). Merely imagining that ‘the item in the jar is poisonous’ was sufficient to move the person to action as if it *were* poisonous (even though they full well knew it was not).

These insights into the imagination can give us parallel insights into ideals. The ideals that we hold, and accordingly which we emotionally respond to, tell us about our values and how we are moved to respond to our values. The contagion feature of the imagination can, thus, explain the ways in which ideals help reconcile us to performing actions that we think are worth

doing, but are inclined not to do. Morality, as well as other standards for behavior, attitudes, and projects, can ask things of us that we don't want to give. As Nicholas Rescher says, "The useful work of an ideal is to serve as a goad to effort by preventing us from resting complacently satisfied with the unhappy compromises demanded by the harsh realities of a difficult world" (Rescher 83).

While contagion is not necessarily harmful, it very well could be—as in the case where we're so moved to act in a certain way by an imagining, that we fail or find it more difficult to take a prudent action. As such, imaginative ideals often may need to be quarantined—recognized as imaginative representations which hold more or less well in the real world.

3.1.3 How Ideals Orient Us

Finally, an important part of the Standard Account is that ideals help "orient our strivings, desires, and practical reasonings" (Coady 51; also see Coady 75, Brownlee 242 and 252). Rescher puts the same point even more poetically: that "Ideals serve to orient and structure our actions and give meaning and significance to our endeavors. They are guiding beacons across the landscape of life—distant, even unreachable points of reference that help us to find our way" (Rescher 119-20).

How does an ideal provide us with orientation? After all, a mere goal ('take out the trash') can also provide us with some kind of orientation. A stipulative ideal ('follow these rules for excellence') could also provide us with orientation. So what is the special sense of 'orientation' that Rescher is describing?

At its best, we know that thinking about ideals can help us to describe and inspire what we should work toward, and serve as a model for conduct now (Hill 3-4). Imaginative ideals—as

opposed to their stipulative counterparts—are effective in a special way because they involve immersion, mirroring, and contagion.

Consider Scanlon’s discussion of the normative ideal of friendships and other relationships:

“It is important to distinguish, here, between the normative ideal of a relationship of a certain kind, such as friendship, and particular relationships of that kind, which hold between particular individuals. The normative ideal of a particular kind of relationship specifies what must be true in order for individuals to have a relationship of this kind, and specifies how individuals in such a relationship should, ideally, behave toward each other, and the attitudes that they should have. It thus sets the standards relative to which particular relationships of this kind exist and the (higher) standards relative to which such relationships can be better or worse, and can be seen as impaired.” (Scanlon, *Moral Dimensions*, 133-34).

Scanlon’s discussion is framed as a stipulative ideal: the facts about a relationship that matter for it to be especially good of its kind. In short, it indicates the standards of excellence for relationships. The stipulative ideal has its role: we can use these standards as a checklist for what we might try to see or bring about in our own friendships and other relationships.

However, we also think about ideals of friendships in terms of what it would *be like* were we to see such friendship in action. Frodo and Sam from *The Lord of the Rings*, or Abileen and Minny from *The Help* illustrate the standards of excellence Scanlon talks about—they bring the ideal of friendship closer to our experience, by allowing us to have perception-like imaginings of what it would be like to see such friends interact (or what it would be like to *be* such a friend). These facts, behaviors, and attitudes are what we imaginatively immerse in (as we do when we read the text, or watch a play or film of the story). And imaginative immersion might have a kind of ‘contagious’ effect on our behavior—by imagining what it would be like to be a good friend (even in the face of hardship), we might be able to do the same for our own friends.

3.2 Our Relationship to Imaginable Ideals

3.2.1 Beliefs about Held Ideals

Even though the imagination is important for understanding ideals, my account has not yet fully characterized the beliefs I hold toward any of my personal ideals. If I hold an ideal, I consider that *ideal object* or *ideal characteristic of an object* to be high or the highest value for that kind of thing. Moreover, there is a difference in whether we refer to something as ‘*an ideal x*’ or ‘*the ideal x*’. If someone considers a representation to be ‘*the ideal x*’, then the agent holds a belief that it *best* represents the valued criteria (under certain parameters) for its species. If it is considered to be ‘*an ideal x*’, then it is amongst the set of objects which best represents those valued criteria.

The specifying of parameters for an ideal is important—it tells us what ‘kind of thing’ the ideal is supposed to capture. I might restrict the ideal to a certain degree of possibility (an ideal of what’s personally possible, or what’s empirically impossible). They might include holding several features of the target object constant and permitting variations elsewhere. Or they might specify a particular *person* or *context* for which the ideal is taken to hold. The parameters tell us what sorts of characteristics the ideal must include (or at least be compatible with) as well as indicate the domain we are concerned with. For example, if I am considering what the ideal career for me would be, I might wish to restrict it to the training I have and who I am and my values now, not the counterfactual case in which I could go back fifteen years and re-train for what might have been a better career for me given what I now know are stable traits of character.¹⁸

¹⁸ Of course, this example opens up questions about *which self’s values* would be acted on. It is common to think “if I knew then what I know now”. Yet my fifteen-years-ago self might not have been able to know otherwise. My current

We might also want to notice that we sometimes disagree with others who might hold an otherwise very similar ideal. Consider, for example, internecine disputes among a political party or a movement. In order to explain this feature of ideals, we need to discuss two kinds of standards for ideals: *inclusion* standards and *fidelity* standards. One way in which our ideal-formation proceeds is by building ideals out from an existing, real-world object: a person, a character trait, a community, or so on. If that is the case, then we will want to talk about *which* characteristics of the object that we should include in the ideal (especially when we have to discard one desirable characteristic that is incompatible with a more-desirable characteristic). These are the *inclusion* standards we hold for an ideal: the criteria of excellence.

Fidelity standards, in contrast, come from the degree to which a characteristic is permitted to depart from the real-world object. This might be especially true when talking about political and moral ideals: can we find an ideal useful, if it departs from what we think is a correct understanding of human psychology—as in the case in which people are taken to be more trusting, or more other-regarding, than they generally are.¹⁹

The paradigmatic ideal includes many high-value criteria that are satisfied that they typically result in a secondary belief about the ideal's *feasibility*. If we recognize that the criteria for excellence are either incompatible or very difficult to jointly satisfy in a real-world object, then we will accord a corresponding amount of confidence in whether or not the ideal is feasible.

As pointed out by Brownlee, we are not talking about deep impossibilities for ideals—for example, a logical impossibility—unless there is a direct conflict between two or more desired

self's values are the product of the experiences I have had. Nonetheless, I also now might be sensitive to various stable traits of character of mine that my past self wasn't aware of, but could have been.

¹⁹ This language ("inclusion" and "fidelity" standards) is borrowed from Weisberg, "Three Kinds of Idealization", a paper on idealization in philosophy of science.

criteria for the object. Instead, we get something generally between a strongly unrealizable ideal (when we think that, for example, an ideal is empirically impossible—as the ‘virtuous person’ might be, depending on one’s commitments about human psychology), and a weakly unrealizable ideal (as when we think that an ideal might be impossible for a certain person or under certain conditions, but it is possible in the rarest of cases—and we might have a real person who serves as an ideal in some way, like the way we take our artistic and athletic idols to be).

The feasibility belief depends upon what we have observed about the world and what we think might happen in the future. I suggest we use the following terminology about ideals: we call those that we think are generally feasible ‘models’ or ‘goals’, and contrast them to ‘ideals’ that we are either certain are unrealizable or are uncertain about their realizability.

3.2.2 Role in our lives

Our distinction between held or personal ideals and recognizable ideals becomes important when we articulate what role ideals should play in our lives. A personal ideal will include a desire that the ideal comes to pass (within certain parameters). If the parameters for an ideal hold (and pursuit of the ideal doesn’t undermine some other, more important plan), I might be moved to action on its behalf. For example, someone might think that his ‘ideal self’ is the kind of person he wants to be in this world,²⁰ and he would be right to try to become more like that ideal. Contrastingly, if those parameters *do not* hold, I may not be moved to direct action—although perhaps I may take on some kinds of indirect action on behalf of that ideal. I might, for example, think that my ideal self ought to be more willing to trust others, even though they

²⁰ The parameters, then, would fit the person’s best assessment of how the world is and how to act in the future. His “ideal” would look to the kind of person he wants to be in his future life, not the kind of person he would want to be if he had different talents or experiences or if he was a different person altogether.

recognize that there are many who take advantage of trust. If I am untrusting, I might work to be more willing to trust those who have established themselves as trustworthy, even if I stay wary of others.

For any recognizable ideal, there are likewise beliefs and desires associated with that ideal. One would need to believe that according to some person, or to one's best understanding, it represents the highly-valuable features of the object or characteristic. However, for merely recognized ideals, the desire condition may vary: the non-holder might approve of the ideal—he might have some sympathies, even though he might not desire that it comes about under the same parameters as the holder. However, he can see why someone *would* share in that desire. For example, an atheist might like and appreciate how a religious person might subscribe to a certain picture of the virtuous person (e.g., trying to live one's life like Jesus would). In other cases, the non-holder might disapprove of the ideal: she might think that the outcome pictured by the holder is undesirable, or undesirable under parameters I take to be more appropriate. For example, a person might recognize that someone they know abides by an ideal of womanhood that is deeply sexist and fails to respect women. They recognize how it has force in their acquaintance's life, but think that it is an ideal that is bad to hold.

Depending on one's valuation, a recognized ideal might motivate me to act. For the approving or sympathetic non-holder, the ideal may justify some limited action even if he does not consider it his own ideal. This may be true especially if it is held by someone he cares about (and it does not require anything that would contradict what he wishes to do). A parent might be supportive of his teenager's turn to vegetarianism, and while not motivated to *be* a vegetarian himself, will support their child's project by no longer buying or cooking meat for the household.

However, when one disapproves of the ideal, it may justify some direct action to oppose what she believes are wrong-headed ideals.

Ideals can play various roles depending on how attached we are to the ideal—and the kind of attachment we have. There are many kinds of action that might be taken on behalf of ideals. We may aspire towards ideals; we might hope for an ideal; we might just appreciate an ideal but not let it take over much of our intellectual energy or practical action. The emotions and desires we hold toward ideals can launch us toward these varying kinds of action, depending upon the parameters we think hold for the given ideal.

Depending on how much we are emotionally involved in the ideal, we might find ourselves spending time on the ideal in various ways. Again, if I *hold an ideal*, then at minimum I must think it would be good for the ideal to come to pass (including desiring that it come to pass under certain conditions). The desire is directed toward something that is an object of importance for me, and the values captured by the ideal must be stably desired by me over time.

The kinds of desires associated with the ideal may result in various kinds of behavior. Consider a few different kinds of attitudes towards ideals: admiration, appreciation, or aspiration. If I merely admire or appreciate an ideal, I might desire that it might come about, but I might not desire that I take action to bring it about. However, I might be more apt to defend the ideal against its critics; acknowledge its importance in others' life, and other minimally-demanding activities.

For more feasible-seeming ideals, we think that the responsible agents are licensed to act more in a way that would bring about the ideal (or at least to get closer *to* the ideal). If I aspire toward an ideal, then I will desire that it might come about, and also desire that I take action to

bring it about (to whatever degree is within my power).²¹ Moreover, the parameters that hold for the ideal must hold for it to make sense for it to be something I aspire to—that is, it must be an ideal that fits me or my world. Otherwise, for me to take direct action toward the ideal, I must be able to think that the parameters/conditions that hold for me are good approximations and thereby can provide good guidance. For example, even though that one might recognize that their political ideal is utopian (and thus impossible), one might also think that it is the kind of ideal that can also be approximated without problematic side effects, and so can be the object of direct striving.

For infeasible ideals that might not be good to approximate (and so which might not provide good direct guidance), I might imaginatively immerse in the goodness of the ideal—and thereby source motivation from it toward more prudent courses of action. For example, a person might find communism appealing but think that real-world communistic experiments have shown its failures. They might regularly think about and appreciate the ideal, but not want it to come to pass in the world. Perhaps they might practice various kinds of interactions with their neighbors that reflect the appeal of contributing to one's capacities and go no further. Similarly, I might merely hope that better conditions for the ideal comes about by some agency beyond my own. This might be the case when I wish for a deity to assure that we all receive our just deserts.

4. The Imagination and its Pathological Uses

Various commentators note that while ideals can be immensely valuable to us, it can sometimes badly misdirect us in guiding our action. Isaiah Berlin counsels us to be wary of

²¹ Compare Strawson 2 and Fischer 215.

ideals, noting how totalitarianism, nationalism, racism, and religious bigotry are often centered on ideals (Berlin 1-2). He writes that “the search for perfection does seem to me a recipe for bloodshed, no better even if it is demanded by the sincerest of idealists, the purest of heart.” (Berlin 15). Likewise, Coady marks these same fears: that ideals are dangerous and delusional (Coady 50).²²

Of course, many of us also think that attentiveness toward ideals has also made us *better* than we might have been otherwise—scientific rationalism, liberal democracy, or spiritual or religious exemplars can be wonderful guideposts and inspirations. So where do ideals go wrong? *Why* do they go wrong? I propose that the account of ideals that I have put forward, with its focus on the function of the imagination, may give us answers.

4.1. Kinds of Pathology

There are many kinds of pathological ideals and pathological uses of ideals, both personal and social or political. Most notably, there are vicious ideals (unjust or inappropriate ones), as already discussed (Coady 63, Brownlee 243; Section 3.2.1). There are, however, other kinds of pathology still worth discussion: we can be delusional about how an ideal can or should be pursued; fanatical pursuit of an ideal can lead to horrifying outcomes; being ‘dreamy’ or ‘having one’s head in the clouds’ can lead to a failure to attend to real-world circumstances around us; being self-satisfied that the world is already exemplifying the ideal can stifle progress and permit injustice. For each of these pathologies, I suggest, a misfire of the imagination is often at root.

²² Dewey writes, on this topic, “What we ardently desire to have thus and so, we tend to believe is already so. Desire has a powerful influence upon intellectual beliefs. Moreover, when conditions are adverse to realization of the objects of our desire—and in the case of significant ideals they are extremely adverse—it is an easy way out to assume that after all they are already embodied in the ultimate structure of what is, and that appearances to the contrary are *merely* appearances. Imagination then merely supervenes and is freed from the responsibility for intervening. Weak natures take to reverie as a refuge as strong ones do to fanaticism. Those who dissent are mourned over by the first class and converted through the use of force by the second.” (Dewey 22).

4.1.1 Delusional and Fanatical Uses of Ideals

Sometimes, we act according to an ideal because we fail to recognize that the ideal, or ways to pursue the ideal, are inappropriate for current circumstances. That is, staying with the demands of the ideal would produce, as Coady says, not just sub-optimal but downright *bad* outcomes (Coady 73). This happens regularly in foreign affairs, he notes (Coady 73)—a country’s pursuit of security might lead to policies of border control that, in turn, lead to grave harm coming to refugees who are turned away. Or our pursuit of democracy might lead to unjust interventions where one country undermines the sovereignty of another. In short, delusional uses of ideals “ignor[e] relevant differences between ideal and real world conditions, urging us to follow policies suitable for ideal conditions but impossible, impractical, dangerous or otherwise inappropriate for us to try to institute or even approximate.” (Hill 2).

This kind of pathology is due to a kind of failure of the emotional imagination: we are more likely to anticipate feeling bad after an undesirable outcome (for example, lack of progress toward a desired ideal). However, our “psychological immune system” is able to take the sting out of negative events: we transform unhappy outcomes into happier effects (Dunn et. al. 335). We might discount the bad outcomes as ‘another step’ toward the ideal, even when they are a consequence of idealistic thinking.

The fanatic’s pathology is similar. It is a pathology in which she has failed to consider the interests of others whose interests are relevant, or to have given enough weight to those interests (although it could include other kinds of failures as well). Fanatics are especially problematic when they are focused on social and political ideals that would require the consent of others. This seems to be a kind of failure to sufficiently imagine what the ideal would be like, or what bringing about the ideal would be like. Those who fail to do so, we tend to accuse of have not

given enough thought to the ideals they purport to hold—while they may be able to details the kind of excellence they think are important, they have not adequately gained insight into their ideal by virtue of failing to be responsive to the mental lives of others.

4.1.2 Dreaminess and Self-Satisfaction

Another kind of misfire is when we get so distracted by wishful thinking that we fail to be receptive to evidence about the world around us. The person who has their ‘head in the clouds’, or is ‘overly optimistic’ on behalf of an ideal has failed to recognize that there is a certain degree of immersion in the imagining beyond which it is no longer good for us. Immersion can distract us from evidence about our other reasons to act, or the probability of the desired outcome. For infeasible ideals especially, it might be pathological to spend too much time on, or too-deeply immerse in the ideal. This overfocus on imagining the ideal distracts from being able to critically think about the parameters that might hold for the ideal, or whether it conflicts with other values—including the needs of ourselves or others.

A related pathology is when we see the world as already embodying the ideal when, in truth, the world is far from it (Coady 74). We might not respond to injustices in the world because we have failed to recognize aspects of reality that don’t correspond to the story we have told ourselves about what is true. While we have various kinds of techniques to try to be sensitive to this tendency—trying to listen to the experiences of those whose lives are unlike our own, being self-critical—this kind of pathology can be very persuasive.

Dreaminess and Self-Satisfaction tend to happen when we dwell on imagining the attainment of the desired outcome. Even if one’s evidence for the probability of the outcome stays the same, that imaginative emphasis on the *aim* leads us to have expectations of the desired outcome (Oettingen and Kappes 396). Imagining also distracts us away from negative feedback

that suggests that the outcome is less probable. Instead, we are better served to mentally contrast the desired future with reflections on reality (Oettingen and Kappes 396). I might recognize that I am quite far from my goal, and contrasting would make me derive meaningful information from that negative feedback, while preserving one's optimistic view (Oettingen and Kappes 409).

4.2 Guidance for Ideals

One of the benefits of an account of ideals as *imaginative* is that understanding how the imagination can lead us astray means that we can identify which ways of holding an ideal are non-viable, and which are. These pathological uses of ideals—I suggest—can be avoided if we are attentive to normative guidance for the imagination.

Contrast Coady's account of why ideals go wrong. He suggests that ideals go wrong when we are paying single-minded attention to them. He recommends, instead, that we must be thoughtful about our other obligations and ideals and how those might point us to a different course of action (Coady 63). If Coady is right, we might think that there are various concerns that have a claim on our attention (tasks, obligations, concerns), and it is up to us to ration out our time and attention appropriately.

However, many of us—when we recognized that we have strayed towards 'idealistic' thinking—don't think it's simply a matter that was a failure of time- or attention-allocation. Instead, we feel *compelled* by the ideal in some way, and it doesn't feel fully under our control to be able to pull ourselves away. Coady's analysis may sometimes be right. However, my account can explain more cases of pathology, and provides an explanation (based on the misfires of the imagination) of what went wrong.

While the account of ideals I have given assesses them as imaginative representations of excellence for an object or characteristic of an object, we have seen that we must be cognizant of various structural features of the imagination. The type of imagining we exhibit about an ideal—and the pathologies of imagination—can have grave implications for well-functioning ideals. Ideals must be situated within certain parameters in order to understand the committed beliefs of a holder. They can play varying roles in our lives depending on how much, and what kind, of emotional attachment we have to them.

Conclusion

What I offered, I hope, is a conception of ideals that is responsive to contemporary work on the function of the imagination. While in many ways it is rooted in the Standard Account, it can do better work than the Standard Account in a few ways. It can better explain our tendency towards imaginatively engaging with our ideals. It also helps to explain how the features of the imagination can explain the characteristic functions of ideals. The functions of mirroring and contagion, and how we immerse in imaginative ideals, illuminate both the characteristic functions of ideals and the pathologies that can result from pursuing ideals.

WORKS CITED

- Audi, Robert. "Wrongs within Rights" in *Normativity (Philosophical Issues, Vol. 15)*, edited by Ernest Sosa and Enrique Villanueva. Boston: Blackwell Publishing, 2005.
- Berlin, Isaiah. *The Proper Study of Mankind*, edited by Henry Hardy and Roger Hausheer, foreword by Noel Annan, and introduction by Roger Hausheer. London: Random House UK, 1997.
- Broad, C. D. *Ethics*, edited by C. Lewy. Boston: Martinus Nijhoff Publishers, 1985.
- Brownlee, Kimberley. "Moral Aspirations and Ideals" in *Utilitas*, Vol. 22, No. 3, September 2010, pp. 241-257.
- Coady, Nicholas. *Messy Morality*. New York: Oxford University Press, 2008.
- Currie, Gregory and Ian Ravenscroft, *Recreative Minds*. New York: Oxford University Press, 2002.
- Damasio, Antonio, Daniel Tranel, and Hanna Damasio, "Somatic Markers and the Guidance of Behavior: Theory and Preliminary Testing" in *Frontal Lobe Function and Dysfunction*, edited by H. Levin, H. M. Eisenberg, and A. L. Benton, Oxford: Oxford University Press, 1991, pp. 217-29.
- Decety, Jean and Jennifer A. Stevens, "Action Representation and Its Role in Social Interaction", Chapter 1 of *The Handbook of Imagination and Mental Simulation*, edited by Keith D. Markman, William M. P. Klein, and Julie A. Suhr. New York: Psychology Press, 2009.
- Dewey, John. *A Common Faith*. New Haven: Yale University Press, 1960. Online: <http://site.ebrary.com.libproxy.lib.unc.edu/lib/uncch/detail.action?docID=10579343>
- Dunn, Elizabeth W., Noah D. Forrin, and Claire E. Ashton-James. "On the Excessive Rationality of the Emotional Imagination: A Two-Systems Account of Affective Forecasts and Experiences" in *The Handbook of Imagination and Mental Simulation*, edited by Keith D. Markman, William M. P. Klein, and Julie A. Suhr. New York: Psychology Press, 2009.
- Emmet, Dorothy. *The Role of the Unrealisable*. New York: St. Martin's Press, 1994.
- Faude-Koivisto, Tanya S., Daniela Wuerz, and Peter M. Gollwitzer. "Implementation Intentions: The Mental Representations and Cognitive Procedures of If-Then Planning", Chapter 5 of *The Handbook of Imagination and Mental Simulation*, edited by Keith D. Markman, William M. P. Klein, and Julie A. Suhr. New York: Psychology Press, 2009.
- Fischer, Jeremy. "Being Proud and Feeling Proud: Character, Emotion, and the Moral Psychology of Personal Ideals" in the *Journal of Value Inquiry*, Vol. 46 (2012), pp. 209-222.

- Gendler, Tamar. *Intuition, Imagination, and Philosophical Methodology*. New York: Oxford University Press, 2010.
- Goldman, Alvin. "Imagination and Simulation in Audience Responses to Fiction" in *The Architecture of the Imagination*, edited by Shaun Nichols. Oxford: Clarendon Press, 2006.
- Hare, R. M. *Freedom and Reason*. New York: Oxford University Press, 1963.
- Hill, Jr., Thomas E. "Is Kantian Ethical Thinking Utopian?" (manuscript)
- Kavka, Gregory. "Moral Ideals". Doctoral Dissertation, University of Michigan for Doctor of Philosophy (Philosophy), 1973.
- Kant, Immanuel. *Groundwork for the Metaphysics of Morals* [G], translated by Arnulf Zweig, edited by Thomas E. Hill Jr. and Arnulf Zweig. New York: Oxford University Press, 2002.
- Mandela, Nelson. "An ideal for which I am prepared to die". Text of Nelson's statement at the opening of his trial on charges of sabotage, Supreme Court of South Africa, Pretoria, April 20, 1964. Online: <https://www.theguardian.com/world/2007/apr/23/nelsonmandela2>
- Myers, Michael W. and Sara D. Hodges. "Making It Up and Making Do: Simulation, Imagination, and Empathic Accuracy", Chapter 19 of *The Handbook of Imagination and Mental Simulation*, edited by Keith D. Markman, William M. P. Klein, and Julie A. Suhr. New York: Psychology Press, 2009.
- Oettingen, Gabriele and Andreas Kappes, "Mental Contrasting of the Future and Reality to Master Negative Feedback", Chapter 26 of *The Handbook of Imagination and Mental Simulation*, edited by Keith D. Markman, William M. P. Klein, and Julie A. Suhr. New York: Psychology Press, 2009.
- Rescher, Nicholas. *Ethical Idealism*. Berkeley: University of California Press, 1987.
- Santayana, George. *Interpretations of Poetry and Religion*. London: Adam and Charles Black, 1900. Online: <https://babel.hathitrust.org/cgi/pt?id=uc2.ark:/13960/t3028sf4m;view=1up;seq=9>
- Scanlon, Thomas. *Moral Dimensions*. Cambridge, MA: Belknap Press of Harvard University Press, 2008. ProQuest ebrary. Web. 25 January 2017.
- Schellenberg, Susanna. "Belief and Desire in Imagination and Immersion", in *The Journal of Philosophy*, Vol. 110, No. 9 (2013), pp. 497-517.
- Strawson, Peter. "Social Morality and Individual Ideal" in *The Journal of the Royal Institute of Philosophy*, Vol. 36, No. 136. (January 1961), pp. 1-17.

Velleman, J. David. "Motivation by ideal" in *Self to Self*. New York: Cambridge University Press, 2006.

Weisberg, Michael. "Three Kinds of Idealization" in *The Journal of Philosophy*, Vol. 104, Issue 12 (2007), pp. 639-659.

CHAPTER 2: THE SUPEREROGATORY, IMPERFECT DUTIES, AND IDEALS

Heroism, extreme devotion to morally valuable causes, generous acts of charity, and (sometimes) forgiveness are all morally good actions, generally speaking. Yet most of us do not claim that it is morally *required* that we perform these actions. Instead, we say that they are actions that would be too demanding for most of us or that are up to our own discretion. Philosophers have sometimes described these actions as fitting into a different moral category, often called the ‘supererogatory’.

J. O. Urmson famously contended in “Saints and Heroes” that an ethical theory is inadequate if it fails to appropriately account for activities that are morally good (sometimes incredibly so) but *not* required. Many philosophers have agreed with Urmson’s conclusion. Kantian ethics (as well as utilitarianism) was one such target for Urmson’s charge: he took Kantian ethical theory to, at best, require revision in order to account for our intuitions about actions that are actions are morally good but not required of us (“SH” 207).

Consider, for example, two real-life cases of persons we might call a saint or a hero. Fannie Lou Hamer was one of the leaders of the Mississippi and national civil rights movements. Over the course of her life, she was forcibly sterilized, lost her work as a sharecropper and had her family’s personal property taken after she attempted to register to vote, was beaten and arrested for participating in a lunch counter protest, and her work as an organizer and a political activist was politically opposed by President Lyndon Johnson himself (PBS). Another case we could consider is Aitzaz Hasan, a 15-year-old Pakistani student, who sacrificed himself to save

1500 students from a suicide bomber who was trying to enter his high school in 2014. He confronted and physically stopped the bomber outside the school. Although the bomber was able to detonate the vest, Hasan was the only other person killed (Masood).

Both Hamer and Hasan acted in morally exemplary ways that most of us would be hard-pressed to replicate because they came at such a high personal cost. Urmson thought, however, that the Kantian account would not be able to say what is extraordinary about actions like Hamer's and Hasan's. Instead, it would just characterize their actions as *dutiful*: fulfilling duties of beneficence or virtue. The Kantian, he said, does not draw a principled distinction between ordinary and extraordinary actions that fulfil duties. At best, the Kantian could simply acknowledge how extraordinarily difficult it would have been for Hamer or Hasan to be beneficent or virtuous in the ways that they were.

Kantian thinkers have had mixed reactions to Urmson's claim. The result has been a longstanding dispute between Kantians about whether Kantian ethics has room for the 'supererogatory', and whether it even should. Some thinkers, including Thomas E. Hill, Jr., have argued that Kantian ethics can make sense of actions that are 'morally good but not required'. Hill explains that Kant's imperfect duties (that is, the duty to set two ends for oneself: the end of self-perfection and the end of the happiness of others) must be sincerely and consistently promoted, but not maximally so (*HWMW* 216). That leaves Kantians room to say that the behaviors of saints and heroes go beyond what is morally required yet have moral worth. Others such as Marcia Baron,²³ disagree with Hill, thinking that the supererogatory is not a category that Kantians neither need nor should want in their moral theory. Importantly, morality aims at *perfection* even with regard to our imperfect duties. The saint and hero, then, come closer to

²³ Other philosophers who make similar claims include Melissa Seymour Fahmy and Jens Timmerman.

completely fulfilling our duties, even if we're right to say that their behaviors aren't (strictly speaking) *required*.

This paper will first fully set out the Urmson argument that a theory of morality is inadequate if it does not encompass the fact that some actions are morally good but not required. Then it will show how Hill and Baron, respectively, have responded to that charge. Their dispute on the supererogatory has long been focused on the parameters of imperfect duty. Both Hill and Baron discuss these characteristic activities (those that are 'morally good, but not required') squarely in terms of imperfect duties. Instead, I will argue that a Kantian account of the supererogatory—or, more precisely, the category of the morally good beyond what is required—also needs to rely on Kant's conception of ideals.²⁴ Kantian ideals carry the spirit of morality in a different, and essential, way than duty alone. The third section focuses on how ideals may allow us to accommodate the features of 'the morally good but not required' better than an account of imperfect duty alone.

²⁴ The term "supererogatory" is used, but not loved, in this literature. The reason for the wariness is because "the supererogatory" comes with the connotations of how it was used in the medieval church doctrine of indulgences, in which extra merit could be exchanged for money to the church, joining the Crusades, or other prescribed activities (Baron "KES" 38-9n33). One kind of controversy in the tradition is related to the sale of indulgences (a kind of moral extra credit) in exchange for a reduced penance. The Protestant tradition—which Kant was raised in—thought that this practice was corrupt (Hill and Cureton 1-2, 3-4). This transactional notion of the "supererogatory"—where supererogatory actions can serve as some kind of 'extra credit'—is not what is under discussion in this paper, nor in the Kantian literature. Instead, Kantians are interested in the sense of the 'supererogatory' as 'morally good to do but not required'. I will be using the term solely in that sense for the remainder of the paper.

1. The charge: Urmson's "Saints and Heroes"

1.1 "Saints and Heroes"

A traditional categorization of the moral worth of our actions is that every action falls into one of three categories. (1) Some actions are *required* of us—good to do, and bad not to do. We are required to respect the rights of others and meet certain obligations: in ordinary cases, we ought to drive in a safe way and keep our promises. (2) Other actions are *merely permissible* or *indifferent*—that is, they are good to do *or* good not to do. Many acceptable human behaviors, both morally and non-morally valuable, fall into this category. Going to a play and drinking coffee are permissible activities, as is choosing to help a friend move (it would be good of you to help, but no one would fault you if you prioritized other things). (3) Still other activities are *forbidden* (e.g., stealing, cheating, killing)—bad to do, and good not to do. This method of categorizing actions means that we have three mutually exclusive categories: dutiful actions, merely permissible actions, and wrong actions (Urmson "SH" 204-5).

Yet Urmson noted that certain actors and actions—moral saints and heroes, and their characteristic behaviors—cannot fit into any of these three categories. Saints and heroes, according to Urmson, do things that ordinary persons are usually unable to do. The saint, Urmson says, performs actions that are far beyond the limits of duty, even though *self-interest or inclination* would lead most people not to do it. The hero performs actions that are far beyond the limits of duty, even when *fear* would lead most people not to do it (Urmson 201).

The saint and hero certainly aren't acting in a *forbidden* way—we don't even have to worry about that prospect. But neither do we want to say that they are either doing what is 'required' or 'merely permissible'. In the first case, it would be odd to say that the moral saint is "simply discharging her duty" when she acts in morally exemplary ways. After all, saintliness

(and heroism) is a level of goodness that ordinary persons are not able to achieve and as a result we would not want to demand that ordinary moral agents be saintly or heroic (Urmson “SH” 200-1). In the second case, we would not want to say that the saint is acting in a way that is merely permissible—such a claim would fail to recognize the incredible moral goodness of her action (Urmson “SH” 201).²⁵

Instead, the saint acts in a way that fits into none of these categories: not the kind of thing that is required of a person, but isn’t ‘merely permissible’. This result, Urmson suggested, shows how a moral theory that *only* categorizes our moral activities in these three ways is inadequate to the “facts of morality” (Urmson “SH” 199). Some interpreters have taken this to mean that Urmson calls for a fourth category of moral acts, that of the supererogatory (see Feinberg 281, for one).²⁶

While Urmson’s essay focused on acts of extraordinary valor and virtue, we might think that ‘supererogation’ refers to a wide variety of activities that can come in lesser and greater degrees of goodness, including doing favors, volunteering, and giving to charity.²⁷ That is, the supererogatory includes both some ordinary behaviors (those that fall beyond what is required of us and are morally good) as well as extraordinary behaviors (those that are well beyond what is required of us, like the actions of the saint and hero).

²⁵ This point is even more obvious if we label the intermediate category as the “morally indifferent”, not the “permissible” (such as in Chisholm, p. 99). Certainly, the activities of the saint or the hero are *not* morally indifferent.

²⁶ Urmson apparently did not agree with introducing a fourth category as a solution to the problem posed by “Saints and Heroes”, and made this clear in later work (Urmson, “HIMT”, p. 161-169, ref from Baron “AKTS” 5).

²⁷ Urmson later suggested that the supererogatory covered a number of types of moral actions that are worthy of distinction from each other as they are from duties and obligations (Urmson “HIMT” 168).

1.2 Urmson's criticisms of utilitarianism and Kantian ethics

Urmson's theory had a significant impact on ethical theory because he charged several popular moral theories as failing to be able to appropriately account for these saintly and heroic actions. These 'inadequate' theories included both utilitarianism and Kantian ethics (Urmson "SH" 207). However, we may think that there are options for both the utilitarian and the Kantian to respond to Urmson's claim.

Urmson truly did pose a problem for the simplest forms of utilitarianism, including some that were prominent at the time that he was writing (Hill and Cureton 2-3). Such simple versions of utilitarianism classify actions as right or wrong (or sometimes indifferent) depending upon whether or not they maximize utility. If action A brings about a greater balance of utility over disutility than all the other available actions (B, C, and D), then it is the right thing to do; performing action B, C, or D would be the wrong thing to do. The reason why I should keep my promise, on one well-known version of utilitarianism, is that even though I would immediately benefit from breaking the promise, it would cause more harm overall (to the promisee, sometimes others, and perhaps even myself in the long term) than benefit if I were to do so.

The activities of saintly or heroic persons often bring an incredible amount of good to the world, even though they may cause some disutility (fear, physical harm, the harms of self-denial) as well. We tend to think that these supererogatory behaviors, perhaps nearly always, bring about more good than all other available actions. That would make them the *right* thing to do, and all lesser actions—helping others only to a moderate degree, for example—would be wrong (though these misbehaviors may be excusable).

Yet only these simple versions of utilitarianism are particularly susceptible to Urmson's objection.²⁸ The more sophisticated utilitarian can say why we ought not demand that moral agents be saintly or heroic much of the time. One well-known option is to borrow from Mill's strategy in *Utilitarianism*, Chapter 5. In that text, Mill suggests that punishments or other kinds of sanctions ought only to be applied to those behaviors that would maximize utility were they to be punished (Mill 48-9). If that is the case, then we can see why it would not maximize the good to sanction any person who fails to be heroic or saintly. There are many people who regularly fail to act in ways that are extraordinarily good and extraordinarily demanding. It would take a remarkable amount of effort, with very little benefit, to sanction or punish them (and doing so without hypocrisy would be difficult as well). Instead, we just praise those who can perform incredible acts and may inspire others to do the same (or at least to go beyond their ordinary habits and practices).

Urmson also charged Kantian ethics with being unable to account for the activities of saints and heroes (Urmson "SH" 206-7), although his interpretation of Kantian ethics is all too brief, and—like the charge against the utilitarian—directed at an overly simple interpretation of the theory. Kantian ethics focuses on actions motivated by the moral law as equally and utterly binding for all rational agents. Importantly, Kant wrote that the only actions that have moral worth are those motivated by the moral law (*G* 4:399-400). If so, it seems that all morally worthy actions *are* morally worthy because they are dutiful, and therefore Kantian ethics leaves no room for morally worthy actions that go 'beyond duty' (Hill and Cureton 4). So we're left with an apparent problem, if we value having an account of the supererogatory in Kantian ethics.

²⁸ Two utilitarian authors who have sometimes been thought to be susceptible to this criticism are Jeremy Bentham and Henry Sidgwick. Yet both of these authors allow for a lot of common-sense morality in their account of decision-procedures, so perhaps a philosopher like William Godwin would be a better target for the criticism. Thanks to Roger Crisp for this point.

2. The Kantian debate

What might a Kantian say, if they were to find Urmson's argument compelling: that we ought to be able to say something about the actions that are morally good but go beyond what is required of us? Kantian ethical theory is almost certainly complex enough to make room for our view that some actions are especially praiseworthy because of their moral goodness, even if they are not required.

Since the publication of "Saints and Heroes", Thomas E. Hill, Jr. and Marcia Baron have been engaged in a steady back and forth about the possibility, and desirability, of fitting these cases of good but not required actions into Kantian ethical theory.²⁹ Hill appeals to how imperfect duties call us to promote certain ends (the happiness of others, or one's own perfection) but gives us latitude in the degree to which we must perform them. Some passages in Kant give us reason to think that we should not be overly rigoristic about these duties, and this gives us an appealing interpretation of Kant that makes room for going 'above and beyond' what morality requires. On Hill's view, Urmson is right to call our attention to activities that are good to do but not required, but Urmson's criticisms of Kant weren't entirely right.

Baron rejects Hill's view. She takes it to be in the spirit of Kant to see morality as aiming at perfection, even with regard to our imperfect duties. When we act like a saint or a hero, we see ourselves as coming closer to completely fulfilling our duties.³⁰ In what follows, I will set out (i)

²⁹ We can identify three interchanges between Hill and Baron: *first*, Hill's (1971) "KIDS" and Baron's (1987) response, "KES" (later published in Baron's *Kantian Ethics, Almost Without Apology*), *second*, Hill's *HWMW* (2002), responded to by Baron and Fahmy's essay in the *The Blackwell Guide to Kant's Ethics* (2009), and *third*, the interchange between the two in the festschrift on Hill's work, *Reason, Value, and Respect* (2015).

³⁰ Richard McCarty, in "The Limits of Kantian Duty, and Beyond" remains unconvinced by both Hill and Baron, and continues to argue the Urmson point: that neither can Kantian ethics accommodate supererogation (like Hill suggests), yet morality does require a category of the supererogatory (unlike what Baron suggests) (McCarty 43).

what Kant says about the distinction between perfect and imperfect duties, and then (ii) Hill's account and (iii) Baron's account in more detail.

2.1 *Perfect and imperfect duties*

The distinction between Kantian perfect and imperfect duties is an important one.³¹ Perfect duties are closely in alignment with the traditional category of 'the required': they include those duties that are strictly prescribed, e.g. to tell the truth, keep a promise, or respect others' rights (*G* 4:421ff). They require us to *never* (or *always*) act in a certain kind of way. Otherwise we would be led to a certain kind of inconsistency: I would fail to treat my and others' rational agency as having the same kind of moral value. Perfect duties are morally required of us, yet they are often easy to discharge, since they generally require refraining from certain kinds of behaviors. (I currently am not killing, cheating, or stealing from anyone!)

Imperfect duties, on the other hand, are duties to adopt and promote certain ends (*MM* 6:389). There are two imperfect duties presented by Kant: one directed towards the happiness of others and the other on perfecting oneself. Whatever one does to promote these two ends must be consistent with meeting one's perfect duties (*MM* 6:389). Thus, I can help a person achieve a life goal of owning a house, but I cannot do so by helping them bilk the seller; I can promote my own skill in playing the piano, but not if I end up neglecting my children. So in contrast to

³¹ Some Kantian duties do entail corresponding rights (Baron "AKTS" 7). I have a right, for example, to demand that you do nothing with my property that I have not consented to. In contrast, the imperfect duties—some cases of charity, for example—promote an important moral end, but the person we could give to cannot *demand* those resources of us (Baron "SKID" 226). Kantian theory divorces the concept of duty from the claims that others can exact from us if we fail to perform an action (contrast Feinberg 278, and Mill's division of imperfect and perfect duties (Chapter 5 of *Utilitarianism*, p. 49-5)). After all, helping others at all, or perfecting ourselves, are both cases in which there isn't a person who can really say that we *owe* it to help them move, or to better our skills at counseling friends in distress.

perfect duties, imperfect duties require us to act *sometimes* or in *some ways* that promote the two mentioned ends, rather than to *never* or *always* act in certain, comparatively precise ways.

The adoption of these two ends is morally required of us, but how exactly we go about promoting those ends is up to our discretion (often described by Kant as having a “wide latitude”).³² I might satisfy my obligation to promote the happiness of others by some combination of assisting my mother to do her taxes, helping my friend move, and volunteering in a soup kitchen—but I could equally do so by cooking for my mother when she’s feeling ill, helping a different friend move, and tutoring high schoolers.

One question of interpretation is to what degree we can be satisfied in how we have promoted these two ends. Hill and Baron’s dispute centers on this topic. What they *do* agree upon is that having adopted the end of others’ happiness and my own perfection, I may not intermittently or infrequently think of them, nor may I merely promote those ends on minor occasions. For instance, I cannot give \$5 to charity every few years, do nothing else other-regarding, and then count myself as having discharged my imperfect duty to promote the happiness of others. After all, Kant recognizes that we are beings with various kinds of capacities and needs (*G* 4:424). We need to promote these ends in order to help to meet others’ needs and develop our capacities. Negligible amounts of self-cultivation and negligible amounts of beneficence are inconsistent with the recognition of persons as having those kinds of capacities and needs.

³² It is important to note what we mean by ‘latitude’—after all, there is a degree of flexibility in how I meet my perfect duties as well. To take the famous example, I might repay a debt by paying in cash or check (or these days, Paypal). Hill gives an excellent explanation of how the flexibility with regard to our imperfect duties is of another kind entirely (Hill *HWMW* 205).

One may ask why these are the only two ends that we are required to promote. We might think that others' perfection, or my own happiness, might deserve a similar status. Kant gives a few reasons to explain why these are not ends that we must directly promote. First, while I can indirectly act in ways that give others resources to promote their own perfection (e.g., providing just institutions, education, opportunities), there is *nothing* I can do to force them to act with a particular kind of motivation. Choosing to do something because it is what the moral law requires is a choice that a person must make for themselves (*MM* 6:386). So while I can assist or support a person in promoting their self-perfection, it would be inappropriate for me to force their hand, and moreover, doing so wouldn't even have the desired effect (that *they* freely set an end to promote their own perfection).

Second, Kant thinks that every person, by virtue of having various empirical needs and desires, will naturally and unavoidably want to secure their own happiness. Duties provide *constraints*, and it would not 'constrain' me to pursue my happiness (*MM* 6:386). Note, however, that Kant does *not* suggest that we deny ourselves happiness. If I act in self-denying ways, I might lead myself to be more tempted to do immoral things. So, indirectly, I should promote my own happiness—but not as a part of duty (*MM* 6:388). Thus, Kant concludes, we have only the two imperfect duties: towards the happiness of others and one's own perfection.

2.2 Hill's account

Hill argues that Kant's ethics is much less rigoristic than is sometimes thought: there "is more room for choice in pursuing moral ideals, and not everything good is required" (Hill "KIDS" 55). Our imperfect duties call us to promote certain ends (the happiness of others, or one's own perfection) but do not specify precisely how, when, or how much. I do not have to

take the opportunity to act in ways that promote some end at *every* opportunity that I am not discharging other duties (“KIDS” 71; *HWMW* 218). At some points, I may choose something that I’d like to do (and is not contrary to perfect duties), but that does not promote one of my imperfect duties. Perhaps I can carve out room in a life full of moral projects to read some pulp fiction or to take up coffee-roasting. So long as all things considered, I have regularly and consistently promoted the two ends of imperfect duty to some degree, then I will be successfully fulfilling my imperfect duties (Hill *HWMW* 204-5).³³

For example, a person might be invested in amplifying the voices and promoting the ends of those in a socially-marginalized population. This might be a choice the person has made as one way to discharge their imperfect duty to promote the happiness of others (within a landscape of other-regarding acts, from cooking dinner for an exhausted friend to counseling students). The person who has chosen this project could do so to a moderate degree: becoming engaged in the movement, participating on occasion, and using their skills and placement to promote that good. She also takes some opportunities to perform activities that promote neither of her imperfect duties (when she permissibly could do so). This would be one way to meet what imperfect duty requires of us.

Yet the person also could engage in the project more robustly—she could work nights and weekends to promote this end, and fill the gaps in her schedule with her remaining other-regarding activities. While she cares for herself as well, she leaves little to no time for simple recreation or time-wasting. While she would be an extreme case of going beyond the reasonable standard, this person may be also behaving in a way that, on Hill’s account, counts as

³³ A different analysis can be found in Heyd’s “Beyond the Call of Duty in Kant’s Ethics”—Heyd instead describes Kant as “torn” between wanting rigorous standards of duty and his interest in ethical ideals, ends, and virtues (Heyd 309).

‘supererogatory’ (even though, ultimately, they are acting in ways that discharge an imperfect duty). Kant’s account of imperfect duty, as interpreted by Hill, “does not specify exactly what or how much a good man will do” (Hill “KIDS” 73, *MM* 6:393). While there is no bright line, the fact that there is room to do morally indifferent things means that there could be actions that go ‘beyond’ what is required of us.

‘Supererogatory’ actions, then, are explained in terms of imperfect duty (Hill “KIDS” 72). They would involve promoting one of the two ends (the happiness of others; my own perfection) from a sense of duty or perhaps for moral reasons, at a time when I am not required to do otherwise by some other duty, and when I have regularly acted to fulfill my imperfect duties (Hill “KIDS” 71).^{34, 35}

Hill argues that we have textual evidence to think that Kant made room for these supererogatory actions. It would be wrong, for example, to take Kant as saying that we must be obsessed with discharging duties. Kant warns us against the person who “allows *nothing to be morally indifferent*”, who “strews all his steps with duties, as with mantraps” (*MM* 6:409). This person’s “concern with petty details, which, were it admitted into the doctrine of virtue, would turn the government of virtue into tyranny” (*MM* 6:409). These passages suggest that Kant thought that there are times in a life in which one could act in a morally valuable way, but one is not obliged to. If the situation is not such that there is another, stringent duty that must be

³⁴ This analysis means that Kantians would not use the phrase “beyond duty” to describe the supererogatory, but that is a difference that we should be comfortable with.

³⁵ There is an objection that Baron levels at Hill which is quite fair (Baron “SKID” 218). This objection points out that Hill’s view shows a kind of infidelity to the traditional conception of the supererogatory. Hill has had to argue that a supererogatory action be motivated by a sense of duty or respect for moral reasons (Hill “KIDS” 71). The traditional supererogationist would say that both the hero who is sympathetic to the needs of those he saves, and the hero who is motivated by duty to assist others, are behaving in a supererogatory manner. The Kantian account of supererogation would only count the second hero as acting in a way that is ‘morally good but not required’. At best, the Kantian could attribute some kind of moral *goodness* but not moral *worth* to the first hero’s action (Hill and Cureton 4, *G* 4:394)).

fulfilled; and that person has “often and continually” acted in ways that promote the ends of imperfect duty, while being motivated to do it *because* it is their duty (Hill “KIDS” 71).³⁶

This interpretation, Hill believes, allows Kantian theory to adequately make sense of the saint’s and the hero’s behaviors—they are promoting their own perfection or the happiness of others beyond the reasonable and consistent amount expected of them. There are places in the Kant corpus—specifically, in the *Vigilantus Lectures on Ethics*—that suggest that Kant also thought that we can find examples of activities that go beyond the “call of duty” (Kant’s examples include a worker who cultivates his talents beyond the call of duty, parents who ensure that their children are not only kept alive but are also educated in a way that leads them to happiness) (*LE Vigilantus* 27:668, 670; original reference thanks to Baron “SKID” [9]n19). In other words, we need not think that Hill has artificially imposed an interpretation on Kant’s ethics, but instead that Kant thinks that there are actions that go beyond what duty requires.

On Hill’s reading, Urmson is entirely right to ask where and how to make sense of actions that are morally good to do but are not required. However, Urmson is wrong to charge the Kantian as not having a way to appropriately explain the actions of the saint and the hero. If Hill is right, then Kantian theory can be appropriately interpreted to make sense of these actions that are meritorious but not required (Hill *HWMW* 234), thanks to his reading of the requirements of imperfect duty.³⁷

³⁶ Hill’s “Meeting Needs and Doing Favors” in *Human Welfare and Moral Worth* (2002) follows from and clarifies this 1971 paper. He first averts a few potential misunderstandings about his account of imperfect duties. He has distinguished imperfect duties from perfect duties insofar as they are directed to the principle of promoting certain ends. However, this commitment is a weighty one: “Kant obviously understood the principle as requiring us to make the happiness of others a serious, major, continually relevant, life-shaping end.” (Hill *HWMW* 206). Nonetheless, there is no mandate to do as much good as possible (Hill *HWMW* 217).

³⁷ In his most recent work, Hill has been careful to say that he is not identifying the ‘supererogatory’ as a *category* of actions—to be precise, Kant’s “scheme for classifying actions leaves room for meritorious acts that have the main

2.3 Baron's criticisms

Baron and Hill agree on several important points. In particular, Baron agrees that Kantian ethics can explain why heroic and saintly actions are, in some manner of speaking, 'good to do but not required' (Baron "KES" 32).³⁸ This is because Kant's sense of 'duty' is not the Millian one in which a duty can be exacted from a person (Baron *KEAWA* 15-16, Baron "SKID" 225, Hill *HWMW* 235). Saintly and heroic actions are dutiful (not "beyond duty") and because they are dutiful they have *moral worth*. Yet they are not *required* because in most cases, one would not need to behave in the saintly or heroic way in order to fulfil one's duties. (It would be a very unusual situation in which one *must* fulfil a duty in that case by behaving in a saintly or heroic way.³⁹)

Nonetheless, Baron is suspicious of having a category of the supererogatory at all. There are a couple of reasons that Baron considers as having the potential to explain why we would want a category of the supererogatory: that we want to demarcate a special status to certain actions, in order to draw out the ways in which they are exemplary (*KEAWA* 48), and that we want basic duties to be within ordinary capacities (*KEAWA* 48). So a category of the supererogatory can be said to provide some practical value.

features of what some philosophers at [Urmson's] time were calling 'supererogatory'" (Hill *HWMW* 234). The original paper was less clear: he wrote that "If Kant's scheme has a place for supererogatory acts, then, they will be found as a subclass of acts which fulfill principles of wider imperfect duty." (Hill "KIDS" 71). Baron, notably, interpreted Hill as recognizing a "class of supererogatory acts" (Baron "KES" 29-30), an interpretation that Hill rejects. Presumably, we should read his work in light of his later clarification.

³⁸ Baron and Hill agree that Kant himself did not explicitly lay out a category of the supererogatory (Hill *HWMW* 236-7). However, the Kantian system to categorize duties is by no means simple, and as such, it does not look very much like the simple tripartite categorization Urmson was considering in "Saints and Heroes" (Baron "KES" 33). This suggests that Kantian ethics may not be as susceptible to the same sorts of concerns.

³⁹ A case that Baron notes where it *would* be morally required to perform a saintly or heroic act is where failing to do so would be a breach of a perfect duty. For example, refusing to sign a false deposition that would send an innocent to his death would be heroic if doing so would put one's life in jeopardy, but nonetheless would be required (Baron "SKID" 216).

Yet Baron thinks that having such a category will *first*, will result in morally undesirable consequences (Baron *KEAWA* 36), and *second*, would require the Kantian to give up an important thesis: that we have a duty to strive to perfect ourselves morally (Baron *KEAWA* 41). However, it doing so she ends up losing the attractive features of Hill's theory in how we should not take Kantian ethics as excessively demanding. In what follows, I will draw out these two arguments.

2.3.1 What goes wrong if we *use* a category of the supererogatory

We might have expected Baron's arguments for why we do not need or want a category of the supererogatory to explain what goes wrong if we *have* a category of the supererogatory. Instead, most of her discussion focuses on the potential consequences: what can go wrong if we *use* a category of the supererogatory. Baron thinks that using a category of the supererogatory might lead us to moral errors of various sorts ("KES" 36ff). In her initial layout, she sets out four varieties of these errors.⁴⁰

The first set of errors are where we wrongly regard or value supererogatory acts compared to acts of ordinary morality. *First*, we might make an Attention Error. We think that supererogatory acts are interesting and compelling, and in paying too such attention to the behaviors of saints and heroes (and others who do more than what is strictly speaking required) we might fail to attend to ordinary morality. A secondary error that could result is that we could make an Importance Error. Kant himself was clearly concerned about such an error. Kant warns us away from using 'noble' examples in moral education (*C2* 5:151, 155; *R* 6:44, 48-9), because

⁴⁰ I have given each error a name in order to help track the various concerns she raised. In the next section, I will add a fifth error from her account.

he feared that “everyday responsibility” would seem to be “petty and insignificant” in contrast (Baron “KES” 36).

Second, we might make a Fungibility Error. The Importance Error led us to think that supererogatory acts have an incredible amount of moral worth—more so than our ordinary moral requirements (Baron “KES” 37). If we make a Fungibility Error, we could wrongly take supererogatory acts as substitutable for our failures in other parts of our moral lives (‘extra credit’) (Baron “KES” 37ff). Baron points out, for example, that we find it easier to offer surplus food or old clothes to the poor as charity than to work to alter the conditions that result in poverty (Baron “KES” 39). If under the influence of the category of the supererogatory we start thinking of ordinary morality as merely optional, or the kind of institution that we can ‘game’, then that would be an objectionable result.

The second pair of errors are ones in which we fail to recognize some truth about our moral selves. These come in two varieties, overestimation or underestimation. If we make an Overestimation Error, we might be disposed to moral fanaticism and exaggerated self-conceit (Baron “KES” 38, also see Kant C2 5:84-5). Accordingly, we may release ourselves from everyday responsibility because we think we are special, or too great for such trivialities (Baron “KES” 37ff). We will not see ourselves as constrained by morality, regarding ourselves as “volunteers”, or that we are constraining ourselves only because our merit (a sort of ‘noblesse oblige’) rather than it being what morality requires of us (Baron “KES” 38). If we make an Underestimation Error, we think that morality is out of our reach. Baron writes, “If moral education emphasizes heroic deeds, morality may seem too remote” (Baron “KES” 37). Those who do not take themselves to be moral exemplars might then take morality to not be ‘for them’ and accordingly fail to hold themselves to moral standards.

It is puzzling that Baron spends so much time on these errors. As Hill points out, her arguments do not show that there *is* no category of the supererogatory, only that such a category could be a bad influence on us (Hill *HWMW* 240). Moreover, there seem to be clear responses that we could make to each set of errors. Our first pair of errors might be solved if we are careful about how we morally educate people, or are cautious about the publicity that we grant to supererogatory acts (Hill *HWMW* 240). Our second pair of errors are matters for how we can be thorough and cautious in our self-scrutiny.

None of these arguments serve as proofs that the supererogatory is not a category worth having or recognizing in moral theory. Now, Baron hasn't mistaken the force of these arguments—she knows that they do not give us reason to 'reject' the supererogatory. However, she thinks that they should make us want to be insistent to hear why we want such a category at all (Baron "KES" 40). I propose that we set these concerns about what could (but does not necessarily) go wrong if we use a category of the supererogatory aside, and hope that her arguments for why we should not *have* a category of the supererogatory are convincing.

2.3.2 What goes wrong if we *have* a category of the supererogatory

Baron has recognized the reasons that many moral theorists want a category of the supererogatory. She has given us reasons to be careful about how we would take such a category to give us guidance for how to act or be a central part of how we morally educate. Yet what reason is there for us to think that something goes wrong if we *have* a category of the supererogatory?

Here she seems to give two arguments. First, she thinks that having a category of the supererogatory inappropriately 'fragments' morality. Such a category separates what I 'have to

do' from what 'I may choose, as I please', and in so doing makes much of morality optional (Baron "KES" 41).⁴¹ The force of this objection is unclear. Separating out what I must necessarily do from what I have discretion to do is *exactly why* moral theorists want a category of the supererogatory in the first place.

Worse, Baron's argument for this point returns merely adds a fifth error to the list that belongs to the previous section: she worries that we would stop thinking critically once we have identified that a certain action would be supererogatory (Baron "KES" 45-47): "The category of the supererogatory seems to me rather to get in the way of moral thinking, especially first-person moral reflection" (Baron "AKTS" 10). Yet again, this is a worry that seems to be better remedied *not* by refusing to have a category of the supererogatory, but in how we use or act on such a category. So we still do not have proof for why we should not have a category of the supererogatory, and must look to her second argument.

Her second argument says that having a category of the supererogatory would be undesirable in Kantian ethics because it would require Kantians to give up the important thesis that we have a duty to perfect ourselves morally (Baron "KES" 41). She says that it is in the spirit of Kant to see morality as aiming at perfection. It is accordingly important to push ourselves to become morally better than we currently are, even with regard to our imperfect duties. However, if we were to hold some actions as supererogatory and others as morally required, we ignore the importance of perfecting ourselves. Instead, when we act in the ways that

⁴¹ Baron, as a result, thinks that Hill is wrong to separate out a category of the supererogatory (Baron "KES" 29). However, Hill has not said that he is proposing a 'category' of the supererogatory. Instead, he simply thinks that we can recognize that in some situations, after careful consideration, we might conclude that performing some action is more than what is required of us (Hill *HWMW* 241).

are morally good but (strictly speaking) not required, we should see ourselves as coming closer to completely fulfilling our duties.

I agree that we should take the duty to morally improve ourselves seriously. Nonetheless, there is a question of how we should take the duty to morally improve ourselves to place demands on us. If we interpret her argument one way, there might be relatively little difference between Baron's and Hill's positions. They both want to resist claims that Kantian ethics constantly demands that one do all that one can (Baron "AKTS" 4, 10; Hill "KIDS" 58-9). So perhaps Baron wants to agree with Hill that Kantian ethics allows room for pursuing non-moral affairs, provided that we have satisfyingly met our imperfect duties in other ways. This would be consistent with Baron's claim that Kantian ethics should not be taken to 'swallow us up' (Baron "SKID" 227-8). Hill is certainly able to claim both that we have room to pursue non-moral affairs *and* that we should work to morally perfect ourselves. Yet even if Baron and Hill agree on the demandingness of Kantian ethics, I suggest that Hill's view of the supererogatory still needs a supplement. I will give this supplement in Section 3.

However, if we interpret her argument a different way, she might be saying that Kantian ethics is *more* demanding than what Hill's account suggests (Baron *KEAWA* 6). Baron takes the duty to morally improve ourselves quite seriously.⁴² She writes that because we have a duty to improve ourselves morally, the imperfect duties "cannot be seen as admitting of a plateau, a point beyond which more conduct of the same sort is supererogatory" (Baron "KES" 42). Even if morality does not require us to maximize, having the maxim of self-perfection means that one cannot simply draw a line around one's "own life" and "own projects" and see morality as

⁴² This is unusual. As will be discussed in the next chapter, other Kantians soft-pedal the demand that we morally perfect ourselves, by instead suggesting that we should 'morally improve' ourselves.

having been satisfied (Baron “KES” 42-43). The famous “ought implies can” principle is, on Baron’s view, turned on its head: instead of thinking that if we cannot do *x*, we have no duty to *x*, we should instead recognize that we are more capable than we think we are (although we are not always honest to ourselves about what we think is ‘too difficult’) (Baron “KES” 44-45).

I do not think that this kind of demandingness is appropriate for Kantian ethics. It seems to be a variation of the ‘maximizing’ morality that Baron takes herself to be opposed to. We can agree that morality *should* make demands on us: we should work to promote our own perfection (even if we’re happy as we are), and that we should press ourselves to perform moral actions (even those we find to be personally difficult). Nonetheless, it seems like her account may be overly demanding. Second, it looks very much like the attitude that Kant warned us against. Remember, Kant has warned us against the person who “allows *nothing to be morally indifferent*”, who “strews all his steps with duties” (*MM* 6:409). On Baron’s account, I must be constantly looking to perfect myself. In Chapter 3, I will give a different—and more satisfying—approach to this duty to morally perfect ourselves.

3. Kantian Supererogation and Ideals

We should think that Hill’s account is appealing in many ways. Hill has said that imperfect duties make demands on us but do not overwhelm us; if we have committed to the two ends and promoted them regularly over the course of our lives, then we may sometimes act in ways that do not promote those ends (so long as perfect duty does not require anything of us, either). The actions of someone who has done far more than fulfils their perfect duties (the Kantian saint or hero, for example) could be called the ‘supererogatory’. His account gives us a

way to think about what is ‘good but not required’ in terms of imperfect duty, and articulates these duties in a way that is not overly rigoristic. However, I contend that a Kantian account of the ‘good but not required’ will do better if it is informed by Kant’s conception of ideals. Our imperfect duties should enter our moral practices as directed toward ideals that we can understand as moving us in special ways (Section 3.1).

In contrast to Hill, I will argue that the characteristic ideals in Kantian thought must be part of an account for how we account for supererogation, because they must be part of how we understand our imperfect duties.⁴³ Ideals reconcile us to our moral project, inspire hope in us, and orient and harmonize our activities. The Kantian ideals are features of Kantian moral theory that direct us beyond the strict obligations of morality, and help us to realize the idea that humanity in each person is worth a certain kind of respectful regard. In what follows, I will survey Kant’s discussion of the importance of ideals and how they play a role in our lives, and then make an argument for why our conception of action beyond duty must include ideals.

3.1 Kant’s conception of ideals

If we focus on imperfect duty, we look to our requirement to adopt two ends: the happiness of others, and one’s own self-perfection. However, I will suggest that we must add in a focus on Kant’s conception of ideals in order to fully characterize what is involved in engaging with the supererogatory. While we are all familiar with Kant’s conception of morality as based

⁴³ Hill mentions our pursuit of moral ideals (“KIDS” 55). However, he doesn’t make much of them, beyond saying that for Kant, “moral worth depends upon one’s disposition to live by whatever demands and ideals are implicit in a rational, moral life.” (Hill “KIDS” 74).

on moral principle (the Categorical Imperative) and the corresponding perfect and imperfect duties, we should not set aside the importance of his corresponding conception of ideals.⁴⁴

Our two imperfect duties are each related to moral ideals discussed by Kant.⁴⁵ We have two imperfect duties—self-perfection and beneficence. The first should be taken as related to the ideal of the virtuous person—what kind of person might we be, if we cultivated our moral and natural talents (*MM* 6:383, *R* 6:122)? The second should be taken as related to the kingdom of ends—what kind of community would we live in, if we reciprocally aided and assisted others in ways that would, overall, promote happiness (as well as good behavior) (*R* 6:131-2, 134, 136)? The interrelation between our imperfect duties and ideals can help us to make sense of the category of the supererogatory.

Ideals do four things for us: they orient and harmonize our activities, they reconcile us to our moral project, and they inspire hope in us. Together, these give us a more sophisticated understanding of both our imperfect duties and how we should conceive of supererogatory actions.

3.1.1 Orientation

First, ideals involve a kind of intention and orientation toward a particular outcome that presents itself as an (out-of-reach) goal for our efforts (*CI* A569/B597). We know that ideals are

⁴⁴ Other Kant scholars mention the importance of ideals for Kantian philosophy. One example is Wood, “Kant’s history of ethics” in *Kant’s Lectures on Ethics: A Critical Guide*, ed. Denis and Sensen (pp. 126-34), who writes that there is “still clearly a place for ideals in modern ethics, in Kant’s view” when speaking about the relationship between principles and ideals, as well as Kant’s assessment of historical moral theories and their relationships to various kinds of ideals (128). I disagree with Wood insofar as he takes Kant to be rejecting an “ethics of ideals” in favor of an “ethics of principles” (part of Wood’s work to show what Kantian ethics can do well that virtue ethics cannot)—it seems more likely to me that he took both parts to be invaluable (135).

⁴⁵ The ideals mentioned in Kant’s texts include the virtuous person, the ideal moral community (the ‘kingdom of ends’), friendship, the highest good (the moral perfection of humanity, combined with their correspondingly deserved happiness), and beauty.

unattainable—they are consistent with human nature, but seldom do we see people even coming close to exemplifying their features (*LE Vigilantus* 27:675).⁴⁶ Yet the adoption of and development toward certain moral ideals is important. In Kant’s words, they are “needful for the elevation of human life” (*LE Vigilantus* 27:675, *MM* 6:469).

Why are ideals needful? Kant seems to think that human psychology is such that our actions must be aimed towards some end (a ‘consequence’) (*R* 6:4). Given that fact about ourselves, if we set certain ends (moral ideals) for ourselves, then we will be directed toward consequences that correspond appropriately to the demands that the moral law makes on us (*R* 6:7n, 6:66). He writes, “since the sensible inclinations of human beings tempt them to ends (the matter of choice) that can be contrary to duty, lawgiving reason can in turn check their influence only by a moral end set up against the ends of inclination” (*MM* 6:380-1, also see *R* 6:6n). So, rather than being as tempted to secure as many goods for *myself* as I can, I might be inspired by a vision of the moral community in which each person has the resources they need to thrive. Rather than spending time doing things that I find entertaining, I might orient myself toward a version of myself who is perfectly virtuous.

Moral ideals set a goal for our efforts, a consequence that we as moral agents could continually approximate toward in a number of smaller actions (*C2* 5:109, 122; *MM* 6:383; *R* 6:61). If we do so, the parts of our psychology that are consequence-oriented (our capacity to feel pleasure, our capacity to desire, for example), will appropriately follow the direction that rationality requires of us. This process involves “awakening a lively feeling of this ideal” (*LE*

⁴⁶ The *Lectures on Ethics*—although troublesome sources, because they were not written or edited by Kant—helps us clarify some of the claims he makes about ideals. For example, a particularly fruitful set of comments can be found in the Vigilantius lecture notes on friendship as an ideal, that is later more or less revisited in the *Metaphysics of Morals*.

Vigilantus 27:681, also see *G* 4:462). So, for example, Kant says “if we consider the gracious consequences that virtue would spread throughout the world, should it gain entry everywhere, then the morally oriented reason (through the imagination) calls sensibility into play.” (*R* 6:24n, also see *G* 4:409). We might think of the same with regard to the ethical community—we think it would be wonderful to be able to live in a community where everyone was treated with the status they deserved, and we were to mutually support each other’s concerns and projects. We may be capable only of approaching our ideals, and not attaining them (*LE* Vigilantus 27:681). Nonetheless, the orientation they provide can help our concern for consequences and our emotions fall in line with what the moral law requires.

3.1.2 Harmonization

Another virtue of ideals is that they permit us to imagine some way to harmonize or make consistent the variety of ends that we aim at in our lives, both for individuals and across community members.⁴⁷

As the previous chapter argued, the imagination can help us to explain the way in which ideals ‘harmonize’ various conflicting features of the object in the world. We know that there are conflicts in desirable features of an object, and to satisfy some to a high degree would rule out others. *Mirroring* is the way in which imaginings can accurately reflect the real-life features of the object we are imagining. As explored in Chapter 1, because ideals mirror, they can allow us to test and adjust those features and to put them into balance. We are regularly stretched between many possible ends, both moral and non-moral—should we be promoting our careers, should we

⁴⁷ Kant makes a distinction between an Idea and an ideal. The former is a concept that is universal (a “guideline”), and the latter, a representation of that concept in an individual (an “exemplar”) (*LE* Mrongrovius 29:604, distinction also described in *CI* A567/B595).

focus on our parenting, have we done enough community service, could we spend time practicing a musical instrument? Often, our difficulties come in deciding which ends to promote at which time.⁴⁸

If we can imagine a harmonious person or community, that ideal may provide guidance for us in how to resolve various impasses in our ordinary decisions. For example, take how we must allocate our time in deciding how to better ourselves. In the *Metaphysics of Morals*, Kant writes that perfection is “taken to mean the harmony of a thing’s properties with an end” (*MM* 6:386). To achieve moral perfection, then, a person would try to cultivate her natural capacities and ensure that they are in line with what duty requires of us (*MM* 6:387, 6:391). We might consider a person who can execute all the human duties (*R* 6:61). It can be helpful to imagine such a person who has managed to strike some choiceworthy balance between not only discharging all their perfect duties, but also promoting the happiness of others and their own perfection in a satisfying way. Or it could be helpful to imagine a person whose characteristics (preferences, strengths, abilities) mirror our own in many ways, and how such a person could be relevantly like the ideal person.⁴⁹

Likewise, a perfect moral community would entail a number of virtuous human beings working harmoniously toward an end (*R* 6:95). Likely, this would be a community with practices and institutions that support people in making virtuous choices, and secondarily, help promote each person’s happiness. But for such a moral community, coordination—or harmonization—is required between persons (*R* 6:139). Our projects can come into conflict and may be mutually

⁴⁸ Hill, in “The Importance of Autonomy”, writes: “While it may be debated whether having a unified personality is in general a moral goal, surely we can agree that it is a morally worthy goal to try to face our important moral decisions with as few as possible of these self-fracturing obstacles.” (Hill “IA” 136-7).

⁴⁹ Chapter 3 will discuss the ideal of moral perfection in more detail.

incompatible. So perhaps an ideal moral community could mirror these features of our ordinary communities, and in so doing help us determine methods to fairly mediate between the permissible projects taken on by individuals.⁵⁰

One apparent problem with ideals providing such guidance is that they look ill-suited for our ordinary lives, because ‘harmonious’ lives and communities are rare in our world and are unlikely to come about. Kant’s insistence that we should cultivate *all* of our capacities (both intellectual and physical) so that we are fit to “realize any ends you might encounter”, however uncertain you are which of them could sometime become yours” (*MM* 6:392) just seems overly demanding. The task seems impossible to achieve, and exhausting to even try to aspire toward.

However, Kant’s ideals may be more personalizable than they seem on the first glance. Kant thinks that there is a degree of choice involved in how we develop our capacities. Kant writes, “Which of these natural perfections should take precedence, and in what proportion one against the other it may be a human being’s duty to himself to make these natural perfections his end, are matters left for him to choose in accordance with his own rational reflection about what sort of life he would like to lead and whether he has the powers necessary for it” (*MM* 6:445). Perhaps a person can use a moral ideal of the virtuous person who resembles themselves—having similar natural talents, cognitive biases and other strengths and weaknesses—but who has managed to strike a balance between cultivating their talents and pursuing projects in a way that leaves them best able to meet their duties, while also living a life that is pleasing to them.⁵¹

⁵⁰ Chapter 4 (and to some degree, Chapter 5) will discuss the ideal of an ethical community in more detail.

⁵¹ This possibility will be explored more fully in Chapter 3.

3.1.3 Reconciliation

These two functions (orientation and harmonization) of Kantian moral ideals can serve to reconcile us to the constraints that morality places on us. As mentioned before, Kant has written that “since the sensible inclinations of human beings tempt them to ends (the matter of choice) that can be contrary to duty, lawgiving reason can in turn check their influence only by a moral end set up against the ends of inclination” (*MM* 6:380-1). Our ideals give us stamina in our practice of morality (*R* 6:61). A pursuit of certain ideals—the morally perfect person, the ideal moral community—may help us shed any temptations to do other than what morality directs us to do. Moreover, the kind of work that an ideal can do to harmonize ends may help us when we are tempted to act in ways that are mutually inconsistent.

We must note that this reconciliation is not inconsistent with Kant’s insistence that moral action must not be motivated by our desires, but instead be motivated by the moral law. The best interpretations of Kant are those that read him as sensitive to our concerns as empirical creatures with needs and wants. We have a number of desires. What matters is whether they are *subordinated* to moral motivation (*R* 6:36, 58). Natural inclinations can still be part of our motivational structure. If our natural inclinations happen to be directed towards a *moral ideal* (one which we find compelling in a number of ways), we might be less distracted by other consequences or emotions that might otherwise tempt us to do other than what morality requires of us.

3.1.4 Hope

Finally, Kant regularly emphasizes the work that ideals can do for us in helping us to hope and to avoid despair. There are two kinds of hope that are at play here: hope for our moral perfection, and hope for our happiness.

First, we must be able to hope that we can become morally perfect.⁵² I may not be able to think of becoming morally perfect as the same kind of ‘attainable’ ideal as I might think of becoming fluent in French, or becoming a lawyer. Yet as Kant writes, a person “must be able to *hope* that, by the exertion of *his own* power, he will attain to the road that leads in that direction, as indicated to him by a fundamentally improved disposition” (*R* 6:51, 6:48). Moreover, we must hope that our past misbehaviors are redeemable (*R* 6:116). Given the ideal, we can trust that even when we are under similar temptations and afflictions, we would again act morally (*R* 6:62). This ‘reasonable hope’ is grounded in that we ordinarily can see steady improvement in our dispositions and one’s improved strength of will (*R* 6:68). It is not certainty (*R* 6:76). It works to “stimulat[e] greater hope and courage and effort in achieving it” (*R* 6:133).

Secondly, we must be able to hope that we will become happy to the degree that is proportional to our desert (*MM* 6:377). Again, Kant is best interpreted as being sensitive to our needs as empirical creatures. If our hope was merely directed toward becoming morally perfect, it would be overly demanding for creatures with needs and wants. However, our hope is directed toward deserving happiness (and receiving it) as well as being virtuous.

The moral ideals give us a way of hoping for our duties, because we can imagine what it would be like if these ideals were to come to pass (thanks to successful execution of our duties). Hope is a balm against temptations to do otherwise than what morality requires. It makes us feel capable of taking on sometimes burdensome tasks, and ensures that we feel like there is a possibility of appropriate reward that will follow.

⁵² Many of us may be wholly unpersuaded that we must hope to become morally *perfect*. Perhaps those readers would prefer to substitute that we must hope to become more *perfect than we are right now*. This is often because we may worry that hoping that we can become morally perfect will overwhelm us. In Chapter 3, I will offer an understanding of this obligation that still directs us to *perfection*, but is not (I hope) in an overwhelming manner.

3.2 Our duties in the direction of ideals

Together, these four features—orientation, harmonization, reconciliation, and hope—show what is special about Kantian ideals. These features then might help us understand what our imperfect duties require of us, and I suggest they are a non-negligible part of doing so. We could, for example, aim to promote the happiness of others or our own perfection without these features. However, then our fulfilment of our imperfect duties would have the ‘thinner’ quality of a kind of rule-following, rather than the thick quality of being informed by an ideal. We can, for example, assist someone simply under the description of ‘doing this would help that person’. Doing so would technically fulfil that duty of beneficence. Yet if we aim to help the person while imagining or ‘envisaging’ that action as directed towards an ideal (say, as a step toward an ethical community) it does an incredible amount of work for us. Our next task is to show with more precision how Kantian ideals are related to our duty.

Kant often uses the language of ‘must’ or ‘duty’ about ideals. He says that we must continually approximate or to ‘elevate’ ourselves to ideals, and that doing so is a duty (*MM* 6:383, 6:409; *R* 6:62).⁵³ We may want to ask whether Kant is introducing new duties—we have perfect and imperfect duties, and perhaps we have a further set of duties towards ideals. Yet that would be an unlikely account—Kant has not given us any indication that he thinks that his standard division of perfect and imperfect duties (*G* 4:421, *MM* 6:390) needs amendment.

One way of thinking about this puzzle is that fulfilling an imperfect duty could be thought of as a ‘stipulative ideal’, each of which has a corresponding ‘imaginative ideal’.⁵⁴

Recall from the last chapter that a *stipulative ideal* is a conceptual assessment of features of

⁵³ With regard to the particular ideal of friendship, he says that friends must “[adopt] this ideal in their disposition toward each other” (*MM* 6:469, also see *R* 6:76 for further ‘adoption’ language).

⁵⁴ See Chapter 1, Section 2.1 for a more detailed discussion of this distinction.

excellence that hold for the target. This might be one way of thinking about what an imperfect duty is: it is to promote a certain moral end. An *imaginative ideal* is an imaginative representation of the highly-valuable features of that object or characteristic we are concerned with. Each stipulative ideal can come with a corresponding imaginative ideal. I take it that the imaginative ideals (most particularly, the virtuous person and moral community) are the corresponding imaginative ideals to the imperfect duties. These come with the features of orientation, harmonization, reconciliation, and hope, and engage with the imagination in ways that, as explored in Chapter 1, can be extremely powerful.⁵⁵ Thus, I take these moral ideals to not generate new duties, but are ways of fulfilling our imperfect duties in a vivid, directed way.

Consider a person who fulfils their perfect duties and promotes their imperfect duties to the minimum degree. This person might be said to be (minimally) meeting their duties. However, it would be odd to think of this person as *also* directing themselves towards an ideal when they have only minimally promoted their imperfect duties. We might think that even if they have volunteered their time at a food pantry, they might have chosen to do so *more earnestly* if they had spent time imagining what it would be like if our community *were* the kind of one where no one went hungry. We might also think that they would have broadened their attentions to more than the food pantry if they had recognized how food insecurity is caught up with the broader reasons for why poverty exists in our broader community.

The way in which ideals orient us is by demanding us to see our imperfect duties as—while ultimately motivated by the moral law—also anticipating an outcome that is beyond the

⁵⁵ Ideals are not without their dangers, something that Kant readily admits. Morality is not *founded* on ideals or derivable from examples, but principles sourced from reason (*G* 4:408-9). We also know that we tend to reformat morality to our own weaknesses (Wood 135)—for example, we think that some degree of helpfulness is ‘good enough’ (even if more is required of us), or mistake our anger as righteous or justified. This might involve portraying the wrong kind of ideal for ourselves. We *also* don’t want to mistake our activities for actual perfection (a kind of delusion) or be tempted into thinking that where we are, morally, is fine (*R* 6:191). This might involve casting ourselves as already ideal.

immediate results of our action. In doing so, it can serve to orient our action beyond mere duty, and help us to harmonize between individual ways of fulfilling our imperfect duties.

3.3 *Supererogation, imperfect duty, and ideals*

While my argument is deeply sympathetic to Hill's view, it goes beyond those expressed in Hill's arguments by saying we cannot locate this category of the "morally good but not required" merely in what our imperfect duties require of us. Hill says that we would do wrong if we fail to have the imperfect duty as an end, and act in ways that develop our commitment to those two ends (our own perfection; others' happiness). We should also think of our action that is 'morally good to do but not required' in terms of the Kantian moral ideals.

3.3.1 Orientation and supererogation

Consider a Kantian hero:

INTERVENING BYSTANDER: We might well think that a bystander, intervening against someone who is prepared to do harm in a public space, may also have an underlying commitment to the importance of respect for persons that fundamentally motivates him to act. He might go knowingly into a situation which might lead to his likely death in order to promote the happiness of others who are at risk of harm if he does not intervene—their continued lives, and the happiness of their family are at stake.

The Intervening Bystander is motivated by and promotes the two ends of imperfect duty in a supererogatory way: far beyond the degree that he must promote those ends.

I will suggest that we fail to act in a supererogatory manner if we are not *oriented* to the ideals associated with those two ends—and the Intervening Bystander should be read as so oriented. Kantian ideals orient us beyond "make others happy" and "perfect yourself": they help us imagine what it would be like to live a rich, complete life where we have cultivated the talents we possess and used them in service of what is right to do, or how it would be if all parties were made (permissibly) happy in coordination with the needs and wants of others.

We can think of ourselves as failing to go ‘beyond what’s required’ if we don’t see ourselves as approximating toward something far beyond our current capacities. Consider the case of a person who has given an extra amount to charity, but does so in an unimaginative and aimless manner (they simply think that ‘it’s their duty; it would be wrong to do otherwise’). While there may be a range of acceptable outcomes they can pursue *by* giving to charity (from ‘a better world’ to ‘a better Detroit’ to ‘benefit for *these* people who I know personally’), how much better if the person is directed toward some vivid consequence, on course to the Kantian ideals. An account of Kantian supererogation focused on ideals will show something different than “I know that it is my duty”: instead, we see “I’m acting in a way that unites with other people to make this a better world (and here is what that would be like)”.

To draw out this claim further, consider a problematic version of the Intervening Bystander:

*INTERVENING BYSTANDER: The bystander, intervening against someone who is prepared to do harm in a public space, may also have an underlying commitment to respect for persons that fundamentally motivates her to act. She survives the encounter (even though her life was at risk). Moreover, she happens to regularly encounter situations that require her courageous character (perhaps she is in a job that requires risking herself for the sake of others, like firefighting). She is fulfilling her imperfect duty to assist others (and, let’s assume, the imperfect duty to perfect herself) far beyond what is reasonably required. Yet when asked why she continues to behave in this way, she answers in an unimaginative though true way: that it is her duty to help others. She reasons that this is a situation in which beneficence is required, and acts accordingly, but she is imaginatively ‘cold’ to what will happen if she does so.

When the *Intervening Bystander helps others, she does so from the right sorts of motivations—that helping others is something required of her by duty—but it comes off as failing to be responsive to the consequences of her actions.⁵⁶ She does not think about where she

⁵⁶ I recognize that some *would* classify the *Intervening Bystander as doing something that is supererogatory. While her behavior may be classified as ‘supererogatory’ by our ordinary use of the term, I think that she is exhibiting a kind

fits into the world, the lives of the people she helps, nor what it means for her character. She is unoriented and unmoored. She is acting far beyond what her imperfect duties require of her, but I would hesitate to call her actions ‘supererogatory’. The Kantian who focuses on ideals to explain the supererogatory can say what seems wrong in this case: the *Intervening Bystander is not oriented to anything beyond the simple facts that someone would be harmed, and she can aid that person. She does not worry about there being a larger story or a larger direction, only the task at hand.

3.3.2 Harmonization and supererogation

Likewise, consider what we can learn from an assessment of a Kantian saint:

CIVIL RIGHTS LEADER: There are numerous cases of state actions that compromise civil rights. Consider a person who may get by—not without hardship, but she may live a life without incredible added harms—without protesting or organizing against those state actions, and it would be understandable to take the easy choice. Nonetheless, she faces up to physical beatings, intimidation, being fired from jobs, ostracization, possible death, and other harms in order to combat these injustices. One central motivation for her actions is that it is deeply important to her that she be the kind of person who will act to ensure that all people (herself and all others) are treated with the respect they deserve. She has taken steps to become a better speaker and motivator, as well as to be courageous enough to stand by her moral convictions in the face of adversity.

The Civil Rights Leader is also motivated by and promote the two ends of imperfect duty: their own perfection and the happiness of others. She has done so in a supererogatory way: far beyond the degree that she must promote those ends. She has also done so, I take it, in a way that makes these ends that we aim at *consistent* or *harmonious*.

One common feature of our ends is that we care about, and aim at, a number of different ends that are practically or otherwise incompatible. We only have so much time, and even have

of behavior that we should find objectionable. As such, I think that a Kantian account of the supererogatory should not permit her behavior to fall within a description of the supererogatory.

trouble deciding how to allocate our time between the projects that we have chosen to put more time towards (our profession, hobbies, family, and friends). As mentioned before, directing oneself toward an ideal can assist in this project: we can think of balanced, whole lives that help us to prioritize certain projects and activities and relationships over other, noticeably good, but ones that may be less important to us.

Assuming that we have taken care of our perfect duties as well, we must mark the danger of becoming overly-focused in our pursuit of certain ends that *do* fulfill imperfect duties, but without attention to how these duties work their way into a full life. A person could be so dedicated to working for a cause that they give insufficient time to their family and friends or themselves. Attention to an ideal of a virtuous person might allow us to balance our moral pursuits and our needs as an empirical person (for self-care, time off, and so on)—and as such, be a wonderfully helpful method for balancing our fulfillment of our duties. Likewise, attention to an ideal of a community of moral agents might allow us to show how the community's single-minded focus on one deserving subset of the community (children, say) might undermine the pursuit to serve the needs of other deserving subset (the homeless, the addicted). Instead of framing our commitments to one set of persons who we have a special relationship to or are passionate about, we might attend to the overall concerns of persons in our community.

Consider a problematic version of the Civil Rights Leader:

***CIVIL RIGHTS LEADER:** This is a person who faces physical beatings, intimidation, being fired from jobs, ostracization, possible death, and other harms in order to combat injustices to a certain group because they are not being appropriately respected as persons, and as a result, their happiness as individuals is compromised. Yet in the course of doing so, he is so concerned with his pursuit (that he be the kind of person who takes steps to ensure that the group that faces injustice is treated with the respect they deserve) that he does not attend to other moral concerns that he should be attending to. He might be neglectful of his family, be less focused on self-care, or ignore the person from the socially privileged group who also needs his assistance. In other words, while he is

deeply devoted to improving himself in a way that will lead to the lessening of a great harm, he is meeting the barest minimums on all other fronts.

The *Civil Rights Leader, like the Civil Rights Leader, may yet be fulfilling his imperfect duties to the degree we think is minimally required. With regard to his self-perfection, he goes far beyond what is required of him—to a supererogatory degree. Yet there is something objectionable about what the *Civil Rights Leader is doing: he is overly focused on his self-perfection in a way that meets his moral obligations, but is unbalanced with the other activities and projects that he otherwise is setting aside (after meeting them to only a minimum degree).

A failure to attend to the Kantian ideals isn't the same kind of wrongness as a failure to fulfill our perfect duties, or failing to fulfill our imperfect duties to an appropriate degree. It involves a failure of orientation and reconciliation that might otherwise improve our commitment to the ends of imperfect duty: the happiness of others and self-perfection. There's something about the ideal of humanity in a person that is inspiring beyond the minimum of "promote happiness to a reasonable degree" or "promote your own perfection to a reasonable degree". It is a moral ideal beyond those bare minimums that are a fuller realization of the idea that humanity in each person is worthy of a certain kind of respectful regard. We can be inspired by the idea that other people like us have all the capacities of humanity: rational and otherwise.

My analysis of supererogation based on ideals has all the benefits of Hill's account, plus more. Ideals help us to structure our commitments to promote the ends of imperfect duty in a special way. We also can be savvy users of ideals, by knowing that they can sometimes lead us astray. Attempts to refine our ideals, and be humble about both how we have portrayed them and how we attempt to approximate them may also be part of practicing morality well.

With this conception of Kantian supererogation in hand, we can see how it allows us to reflect on the Categorical Imperative in a way that benefits us. After all, if the idea of humanity as an end in itself has its full effect in me, then I will make (so far as possible) the ends of others my own ends (*G* 4:430). The bare minimum requirement I must fulfill is to avoid treating people as mere means, and adopt the principle that others' happiness and their ends matters too. That minimum needs to be supplemented by arguments about particular cases (e.g., why should I help this person right *now*). But we also need the ideal of promoting the development of humanity of other people. It will provide orientation to my activities, consistency in the particular ends I choose, and additional motivation when I am tempted to do otherwise.

Conclusion

There are a number of varieties of supererogation. On one hand, easygoing supererogationists might think that it is simply a good thing if we push past the boundaries of what is required of us—every little bit helps—but it's fine if we wanted to play tennis or the oboe (in response to Susan Wolf's argument against the pursuit of moral sainthood). On the other extreme, some moral theories (Peter Singer's version of utilitarianism is the classic case) suggest there may be no upper bound to what is required of us, and thus supererogation does not have a place in moral theory for a very different reason. This question about the demandingness of what morality requires is a classic issue, with adherents on either side. Those who resist the demandingness of morality are generally going to want a category of supererogation.

The Kantian position that I've articulated has some robust points, compared to both of those opposing sides. Like Hill's view, it respects the position of the opponent to Singer—that

morality shouldn't be so outrageously demanding. After all, we only are required to meet our imperfect duties to a certain reasonable degree, and most of our perfect duties are easy to discharge (currently, I'm refraining from killing, cheating, *and* stealing!)

At the same time, we can respect Singer's claim that perhaps morality does demand quite a lot of us—but in a very different sense than he suggests. The fact that we are to pursue ideals means that we can work towards what morality requires of us, even if it's more of a matter of orientation and intention. I will continue this discussion in the next chapter.

The flexibility in how we orient ourselves toward ideals, and to what degree we approximate to those ideals, means that we aren't subject to Wolf's criticisms in the same way the version of Kantian ethics she discusses would be. We are not demanded to continue to give more and more to pursue moral ends. Instead, we are asked to thoughtfully orient ourselves toward those ends, and think about how we can harmonize the many things we care about (moral and non-moral) into a single life.

WORKS CITED

- Baron, Marcia. "A Kantian Take on the Supererogatory" ["AKTS"] in *Journal of Applied Philosophy*, Vol. 22, No. 4 (November 2016), pp. 347-362.
- Baron, Marcia. *Kantian Ethics, Almost Without Apology*. Ithaca: Cornell University Press, 1995.
- Baron, Marcia. "Kantian Ethics and the Supererogatory" ["KES"] in *Kantian Ethics, Almost Without Apology*. Ithaca: Cornell University Press, 1995, pp. 21-58.
- Baron, Marcia. "Minimal Morality, Moral Excellence, and the Supererogatory" ["MM"] in *Kantian Ethics, Almost Without Apology*. Ithaca: Cornell University Press, 1995, pp. 59-81.
- Baron, Marcia. "The Supererogatory and Kant's Imperfect Duties" ["SKID"] in *Reason, Value, and Respect*, edited by Mark Timmons and Robert N. Johnson. New York: Oxford University Press, 2015, pp. 215-231.
- Baron, Marcia and Melissa Seymour Fahmy. "Beneficence and Other Duties of Love" in *The Blackwell Guide to Kant's Ethics*, edited by Thomas E. Hill, Jr. Wiley-Blackwell, 2009. Ebook. pp. 211-228.
- Chisholm, Roderick M. "Supererogation and Offence" in *Brentano and Meinong Studies*. Atlantic Highlands, NJ: Humanities Press, 1982. (First published in *Ratio*, Vol. 5, 1963, pp. 1-14).
- Feinberg, "Supererogation and Rules" in *Ethics*, Vol. 71, Issue 4 (1961), pp. 276-88.
- Heyd, David. "Beyond the Call of Duty in Kant's Ethics" in *Kant-Studien*, Vol. 71, Issue 1 (1980), pp. 308-325.
- Hill, Jr., Thomas E. "The Importance of Autonomy" in *Women and Moral Theory*. Totowa, NJ: Rowman & Littlefield, 1987.
- Hill, Jr., Thomas E. "Kant on Imperfect Duty and Supererogation" ["KIDS"] in *Kant-Studien*, Vol. 62, Issue 1 (1971), pp. 55-76.
- Hill, Jr., Thomas E. "Meeting needs and doing favors" in *Human Welfare and Moral Worth [HWMW]*. Oxford: Clarendon Press, 2002.
- Hill, Jr., Thomas E. "Looking Back: Main Themes and Appreciation" in *Reason, Value, and Respect*, edited by Mark Timmons and Robert N. Johnson. New York: Oxford University Press, 2015, pp. 269-296.
- Hill, Jr., Thomas E. and Adam Cureton, "Supererogation" in *The International Encyclopedia of Ethics*, edited by Hugh LaFollette. Blackwell Publishing, 2013. Published Online February 1, 2013. Print pages pp. 5070-5078.

- Kant, Immanuel. *Critique of Pure Reason* [C1], translated by Norman Kemp Smith. London: Palgrave Macmillan, 2007.
- Kant, Immanuel. *Critique of Practical Reason* [C2], translated by Lewis White Beck. New York: Macmillan Publishing Company, 1956.
- Kant, Immanuel. *Groundwork for the Metaphysics of Morals* [G], translated by Arnulf Zweig, edited by Thomas E. Hill Jr. and Arnulf Zweig. New York: Oxford University Press, 2002.
- Kant, Immanuel. *Lectures on Ethics* [LE], edited by Peter Heath and J.B. Schneewind, translated by Peter Heath. New York: Cambridge University Press, 2001.
- Kant, Immanuel. *Metaphysics of Morals* [MM], translated and edited by Mary Gregor. New York: Cambridge University Press, 1996.
- Kant, Immanuel. *Religion within the Boundaries of Mere Reason* [R], edited by Allen Wood and George di Giovanni. New York: Cambridge University Press, 1998.
- Masood, Salman. "Call to Honor Pakistani Boy in Fatal Tackle of a Bomber" in *The New York Times*, published January 9, 2014. Accessed May 15, 2016. Online: http://www.nytimes.com/2014/01/10/world/asia/call-to-honor-pakistani-boy-in-fatal-tackle-of-a-bomber.html?_r=0
- McCarty, Richard. "The limits of Kantian duty, and beyond" in *American Philosophical Quarterly*, Vol. 26, No. 1 (January 1989), pp. 42-53.
- Mill, J. S. *Utilitarianism* (2nd Edition), edited by George Sher. Indianapolis: Hackett Publishing Company, 2001
- PBS, "Biography: Fannie Lou Hamer". Accessed May 15, 2016. Online: <http://www.pbs.org/wgbh/americalexperience/features/biography/freedomssummer-hamer/>
- Singer, Peter. *Practical Ethics*. New York: Cambridge University Press, 2011.
- Strawson, Peter F. "Social Morality and Individual Ideal" in *Philosophy*, Vol. 36, No. 136, (1961), pp. 1-17.
- Timmermann, Jens. "Good but not required? Assessing the demands of Kantian ethics" in *Journal of Moral Philosophy*, Vol. 2, No. 1 (2005), pp. 9-27.
- Urmson, J. O. "Saints and Heroes" in *Essays in Moral Philosophy*, edited by A. I. Melden. Seattle: University of Washington Press, 1958.
- Urmson, J. O. "Hare on intuitive moral thinking" in *Hare and Critics*, edited by Douglas Seanor and N. Fotion. New York: Oxford University Press, 1988.

Wolf, Susan. "Moral Saints" in *The Variety of Values: Essays on Morality, Meaning, and Love*. Oxford Scholarship Online (Published Online: December 2014). New York: Oxford University Press, 2015. Accessed April 17, 2017.

Wood, Allen. "Kant's history of ethics" in *Kant's Lectures on ethics: A Critical Guide*, edited by Lara Denis and Oliver Sensen. New York: Cambridge University Press, 2015, pp. 126-34.

CHAPTER 3: SELF-PERFECTION, SELF-IMPROVEMENT, AND THE KANTIAN VIRTUOUS AGENT

The slogan “be all that you can be” is inspiring, even compelling. It demands that we look to the capacities we have and secure them to their furthest limit. We feel especially answerable to the claim that we have an obligation to “be all that we can be” morally.⁵⁷ Even if we won’t become the most successful writer, we must at least become as honest and conscientious as we can. The moral dimension of the slogan will be the focus of this paper.

Nonetheless, we realize that we cannot “be all that we can be” in the eighty-odd years that we might hope to live, if we take that to imply moral perfection. We are flawed, morally and otherwise, and we have a limited ability to avoid mistakes and to become better than we are. Our evidence suggests that no matter how much we might want to ‘be all that we can be’, we cannot fully do so.

These observations set up a puzzle for us. There are three claims that are each convincing but are mutually inconsistent: (1) that you ought to “be all that you can be”, (2) that it is not possible to be all that you can be⁵⁸, and (3) that ought implies can. How can we avoid inconsistency without denying something that is apparently true?

⁵⁷ This is not a moral principle that people always accept, but Kantians often take it seriously because Kant accepted a version of it. However, that it served as a US Army motto also suggests that it might have some uptake beyond Kantian ethics. Virtue ethicists may also find the slogan compelling, as well as some (though by no means all) consequentialists.

⁵⁸ Obviously, something should be said about how this seems paradoxical on its face. If I only become so good over the course of my lifetime (after a good amount of effort), surely that’s my ‘being all that I can be’? I suppose that the sense of ‘possibility’ that is implied by this claim indicates not only what deterministically could or will happen over the course of my lifetime, but a kind of physical possibility that would attain only in the closest possible worlds (where

This paper, after setting the problem up in more detail in Section 1, will look to the solutions that have been offered within Kantian ethics. In Section 2, we see how Kant rejects claim (2). Others think that all we need to do is to “be more than you have been”, not the more demanding “be all that you can be”. So in Section 3, I look to how some Kantian thinkers reject claim (1). However, I argue that neither the rejection of (1) or (2) is intellectually satisfying, and that we have good reason not to reject (3). In Section 4, I will show how the moral imagination holds the key to the apparent puzzle raised by our obligation to be all that we can be.

1. The Problem as it Arises in Kant’s Moral Philosophy

The problem of resolving the inconsistency between the three claims is relevant for those moral theorists who feel answerable to the claim that we ought to be the best that we can be, recognize that doing so is impossible, *and* think that ought implies can.

The second claim is quite plausible: that we can at no point be all that we can be. Moral perfection is something we are confident is out of our grasp, even if we can imagine ourselves being morally perfect. It is easy to recognize all the ways in which we are imperfect or limited. Yet those limitations also show where improvement or perfection might be possible—and we feel that such possibilities would be a very good thing. As a result, we feel the force of the demand to use our moral capacities to their fullest extent. The Bible, for example, calls us to “be perfect” (Matthew 5:48). Kant was concerned with a version of this problem (*R* 6:61). He was dedicated to claim (1): that we should “be all that we can be” as moral agents. So he responded

I and my circumstances are relevantly like they are now). This seems to capture what is involved in deathbed wishes where we had hoped to have changed certain things about our lives.

to the puzzle by denying (2), even though it is what most of us think is the *most* plausible claim. Kantians, as a result, have spent a great deal of time re-examining and responding to this puzzle as a result.

Kant's moral views are familiar, yet an exposition may help us understand in what sense and why he thought we must 'be all that we could be'. Agents such as ourselves, to act fully rationally, would act on objective and necessary principles captured by the various formulations of the Categorical Imperative. We can be obligated by practical reason to act on these principles—that is, they constrain us—even while our actions on behalf of the moral law are **autonomous** (i.e., we freely govern ourselves according to universal laws that we can set for ourselves) (*G* 4:447-8). This is because of a characteristic feature of rational agency: we are able to freely set laws for ourselves. If we *choose* to abide by the moral law, then we can be governed by the law yet not give up our claims to be free agents (*G* 4:400). Even though an agent may be tempted to act otherwise, she has a **good will** when she has a fundamental commitment or firm intention to do her duty *because* it is what she ought to do (*MM* 6:379-80).⁵⁹

A person who has a good will (a fundamental commitment to do her duty *because* it is her duty), however, may have a weak will. A **virtuous person** has a will that is not only good but is also strong—making her capable of resisting other, countervailing motivations that threaten to override her moral motivation (*MM* 6:380, 6:405; *LE* Mrongrovius 29:627, Hill and Zweig 41).⁶⁰ Ordinary agents are subject to an internal conflict between our motivation to do our duty and our motivation to act in ways that would fulfill our ordinary wants and desires. To

⁵⁹ We might be curious about whether one could use strength of soul for a negative end—i.e., I have a strong will to do ill. Kant seems to think not; that crimes are caused by the force of inclination (*MM* 6:384).

⁶⁰ The source of an agent's strength of will to fulfill their duty comes from their capacity to *reason* (*MM* 6:405). Reason helps us overcome contra-moral motivations that tempt us away from our duty. It does so by helping us to remember our commitment to prioritizing morality over these other ends (*LE* 245-6).

become virtuous—which Kant says we ought to do—we must cultivate our will so we can abide by our fundamental commitment to morality.

Thus, the Kantian version of “be all that you can be” is directed towards recognizing that as rational agents, each of us has the capacity to have both a good and strong will. Kant tells us that our responsibility is to become virtuous, a version of claim (1). He also believed claim (3), that ought implies can (*R* 6:50).⁶¹ And that left him with a problem. After all, he recognized all the weight of the evidence that we should keep premise (2).

Kant discussed the evidence supporting claim (2) in two ways. *First*, our moral imperfection means that in at least some (and perhaps many) instances, we in fact will not do our duty. Kant acknowledged that we know that we are unable to achieve moral perfection in a human lifetime. Our everyday experiences prove that we are imperfect beings: we are impatient, unkind, lazy, stupid, rude, obnoxious, and so on. Even those who usually conduct themselves well relapse into less admirable behaviors. It is hard to reconcile ourselves to the possibility of always doing the right thing—our track records are already marred, and we don’t anticipate that we will become perfectly good people (*R* 6:29). So if we don’t think moral perfection is possible for us, it means that in some of the cases in which we have tried or will try to do our duty, we will be incapable of doing so.⁶²

Second, our lack of transparency to ourselves casts doubt on our motivations. Kant also noted that even in cases where we think that we are doing the right thing, we could be misled about our own intentions. As he writes, “the depths of the human heart are unfathomable” (*MM* 6:447; see also *MM* 6:393, *G* 4:406; *CI* A551/B579; *R* 6:51). We might have convinced

⁶¹ Of course, contemporary discussions of ‘ought implies can’ suggest an incredible amount of complexity to how we should understand this topic.

⁶² See *MM* 6:446 for Kant’s statement of this idea.

ourselves that we have assisted someone because it was our duty—but in truth, we may have been motivated by our own advantage.

Yet if moral perfection is impossible, *period*, Kant cautions that we are in trouble. We're performing a fool's errand in trying to always do our duty over the course of our lifetimes.

Kant's argument goes something like this: rational agency requires that if we intend to act on an end, it should be possible to realize this end. We have a duty to do our duty at every instance. If moral perfection were impossible, then the moral law would be directed at something that is impossible. Morality then would be setting us a duty to promote a state of affairs that our experiences in the world suggest is a mere fantasy—an “empty imaginary end” (C2 5:114; see also C1 A813/B841, R 6:47-8). It would thus cast doubt on what we intend to do—to do our duty—and make us suspect that the moral law is false (C2 5:113).

Of course, simply because I ought to do something each time, that does not mean that I ought to do something every time. A basketball player may feel she should make a basket every time she shoots, but reasonably not take herself to ought to have made every basket she's ever attempted. Morality, however, is a different kind of project than basketball. Its demand hits us both distributively *and* collectively. This is obvious for our perfect duties, when if we fail we wrong someone. It is less obvious for our imperfect duties, but as Baron and similar thinkers in the last chapter prove, we should take ourselves as needing to push ourselves to be beneficent (and so on), even if it may not be morally required to do so in a particular instance.⁶³

So Kant needed a way to reject (2), despite all the evidence for it. Kant thinks that we do not have empirical reason to believe that we will ever be able to be morally perfect (R 6:67). Yet

⁶³ In other words, persisting with claim (1) should speak to readers who had the intuition that I dismissed Baron's view too quickly in the last chapter (see especially Section 3.2) by articulating that it was too demanding. I take this chapter as a way to grapple with Baron's views about how we should regard our imperfect duties.

we need to believe that we have this capacity. If we cannot achieve the task we are set, Kant says, “the unavoidable consequence of a rational estimate of [the moral worth of] our natural state is a feeling of hopelessness” (*R* 6:71; also see 6:184-5). His solution, however, is controversial. Kant answers this problem by saying that *moral faith* can certify our confidence that we can do what is right and that we have moral worth (*R* 6:45, 6:71-2).⁶⁴

2. Kant’s Solution

Kant’s controversial solution to the puzzle was to use moral faith to help us believe that we could fulfil our duty to become morally perfect—in short, he denies the plausible claim (2) that at no time we can ‘be all that we can be’. If we were to have an immortal soul and the divine assistance of God, Kant says, we could ensure that we could be virtuous. A person must be able to hope that he, by his own actions, will continue to morally improve himself *until* he becomes perfect (*R* 6:51). Given enough time (over an eternal life, say), continual moral improvement could lead to moral perfection (*C2* 5:122). Yet the achievement of perfection through continuous improvement is only possible, Kant thinks, if we possess immortal souls (*C2* 5:122). In an immortal life, the process of gradual self-improvement would move us towards the limit of human goodness: virtue.⁶⁵

⁶⁴ Kant thinks that moral religion is the only way that virtue can be ‘crowned’ with “the hope of the final success of all our good ends” (*R* 6:185).

⁶⁵ It should also be noted that Kant’s discussion of having an ‘immortal soul’ as a solution for the concerns we ought to have about moral despair with regard to our self-perfection may not be philosophically sound, and Kant himself seemed to equivocate about whether he had the correct solution (see Wood, Chapter 5, and especially 182n). Nonetheless, even if we do not fill in an *immortal soul*, we still need to explain what’s going on.

Yet even an immortal soul would not be enough to make ‘continuous improvement over an eternal life’ equivalent to the actual achievement of moral perfection. At each part of our (immortal) lives, we are morally inadequate. So moral perfection also requires assistance from God—more specifically, God’s grace. Kant writes, “we can think of the infinite progression of the good through conformity to the law as being judged by [God]... to be a perfected whole” (*R* 6:68).⁶⁶

We don’t have evidence for the existence of an immortal soul or the existence of God, but neither do we have evidence against these possibilities (*C2* 5:114ff). If these postulates would allow us to conceive of how we could do our duty—and it is important for us to be able to do our duty—then we have good reason to come to believe that they are true.

Failing to have this moral faith, Kant thinks, would lead us to despair. In the Third Critique, Kant considers the case of a morally good man who is not willing to commit to the features of faith: the man believes that there is no God (and presumably, that we do not have immortal souls). This person may try to bring about the highest good (complete virtue and proportionate happiness) (*C3* 5:452). However, Kant assures us, this man is doomed to fail—and must acknowledge that he is doomed to fail (*C3* 5:452-3).⁶⁷

Jacqueline Mariña, in her work, describes the work of the good person who aims to acquire virtue without the buttress of faith in this way: “It is to continue to engage in a battle towards the acquisition of virtue, the inevitable outcome of which is known beforehand; as such, it is to display the courage of a fool” (Mariña 354-5). We know virtue to be in itself impossible

⁶⁶ Modern thinkers might think of this move as ‘taking the limit’ of our moral goodness.

⁶⁷ Kant writes: “The end, therefore, which this well-intentioned person had and should have had before his eyes in his conformity to the moral law, he would certainly have to give up as impossible; or, ... assume the existence of a **moral** author of the world, i.e., of God” (*C3* 5:452-3)

for us; if only virtue will do, then striving after it is irrational when we have no hope of success. Perhaps we can persist for a time toward trying to bring about virtue; perhaps we could find some solace in the marginal steps we take toward virtue. However, when we inspected what we were committed to, we would lose our ability to hope that the complete goal is achievable.

I suggest that we should take the task of directing ourselves to try to ‘be perfect’ seriously. If the Kantian wishes to take the pursuit of moral perfection as our task, Kant’s solution—rejecting claim (2)—has obvious drawbacks. First, many Kantians wish to step away from Kant’s more controversial metaphysical claims. Committing oneself to the existence of God and an immortal soul is a bold claim indeed. Second, these theological commitments may not speak to the secularly-minded person who is looking to Kantian ethics for guidance. This paper will explore Kantians’ secular alternatives to the problem that we are charged with doing something that seems impossible to do.

3. Alternative Kantian Solutions

What should Kantians do, if they wish to avoid the use of a moral faith that depends on controversial metaphysical claims? In this section, I will explore two options: the Self-Improvement View and the Faith in Humanity View. Both of these are variants on the idea that we could reject premise (1). Each, in its own way, says that we are not answerable to the obligation to “be all that you can be”. Instead, the views weaken the claim to say, instead, that each of us is merely tasked with being better than we are right now.

The Self-Improvement View disregards the idea that we must be focused on *perfecting* ourselves—and says that our task as Kantian agents is “strive toward virtue”, not “be perfect”.

As such, its focus is on moral improvement. The Faith in Humanity View is to adopt a *secular* faith in humanity's potential to become perfect, in contrast to Kant's insistence that I have faith that *I* can become perfect.⁶⁸ I think that proponents of these strategies are right to want to avoid despair without an appeal to (non-secular) faith. However, we should find these views unsatisfying for other reasons.

3.1 Self-Improvement View

Many readers think that Kant's response to our puzzle—that an immortal soul is required to avert moral despair—is somewhat overwrought. Even without a guarantee of success, striving does not seem to undermine the integrity of the moral project of fulfilling our duties from duty in all instances. Perfect virtue would remain an ideal that regulates our behavior, but is not the intended end of our action. We might think that there are plenty of reasons that we can instead think of ourselves as morally improving. We might not be successful in the moral project, but we can take solace in how we are doing better than we formerly were.

We know from Kant that virtue is an unattainable ideal for human beings on earth, and these theorists take seriously those passages in Kant where he says that *approximation* to the ideal is what we should be most concerned with. After all, such approximation is what we can expect to achieve at our best (*MM* 6:409, 433n; *C3* 5:469). We orient our lives around virtue and take measures to bring it into existence in some way in some parts of our lives, but we accept that the task is not completable.

We may not be certain about any human's ability to ever be virtuous, much less our own—but getting better? That seems achievable for many people, including, I may hope, for

⁶⁸ The Faith in Humanity View could be either compatible with or independent from the Self-Improvement View.

myself. Lara Denis makes such a point, writing that, “It should take a lot more discouragement to make an agent see morality as fraudulent or irrational” (Denis “KCA” 216). We are still capable of virtue in a satisfying way, even if it isn’t impossibly-high moral perfection.

On Denis’ view, when ordinary agents act with morally good intentions, we honor and foster rational nature in ourselves and others. Even if we are only *oriented* toward the highest good—not focused on its attainment—we still may be able to express respect for rational nature (Denis “KCA” 217). Mavis Biss agrees that we can make modest progress toward the ideal of virtue (Biss 19). She too is unconcerned that striving for self-improvement will get us no further than that. Perfection is an ideal to be approximated toward, not a goal to be achieved. All one must do is to continue to progress toward being more virtuous than one is already (*MM* 6:446). We can make ourselves morally better in all sorts of ways—more sympathetic at the right times, less jealous, better at knowing our reasons for acting in certain ways.

We should not despair because we are achieving something that is worthwhile. Denis writes that “As long as an agent is able, with some regularity, to realize her morally good intentions in the world, she has sufficient grounds for taking the moral law seriously” (Denis “KCA” 216). We shouldn’t despair because we usually make things better than if we hadn’t tried at all (Denis “KCA” 213). Biss says that we ought to focus on our capability *to* resist temptation in many cases, and thereby certify our confidence in our moral agency (Biss 16). Thus even within an ordinary agent’s practice of morality, that person can exercise their ability to act morally while faced with contra-moral motivation (Biss 17). Thus, we do not have to worry about achieving virtue even when we have real suspicions about whether moral perfection is even possible for us.

So on Denis and Biss's accounts, all we need is to believe that we are capable of improvement and to act on that idea in meaningful ways in our lives. We *don't* need to believe that we are capable of reaching perfection. We might think that we should aim to "be all that you can be", but act on what is entailed by that claim: that we should "be more than you have been". Since being more than you have been is actually an achievable project, these authors take it as the more helpful and reasonable method to direct ourselves. Thus, they reject claim (1) as a way to avoid inconsistency.

3.2 Response to the Self-Improvement View

While the Self-Improvement View offers a way out of the puzzle, its solution comes at a cost. Specifically, the view is unable to capture the intuition that we have to "be all that we can be".⁶⁹ If we were to think that our activities as moral agents were like other projects in our lives, it would not be a problem if I neither am a perfect person nor have a hope to reach perfection. The musician can still enjoy playing and think her project of practicing and performing is worthwhile, even though she knows that she will never fully master her instrument.

The Self-Improvement theorist would suggest that the same is true about our drive toward being morally perfect. A person may understand that they are a non-perfect and non-perfectable person, but with unknown limits to perfect themselves. Yet it looks like plenty of people are motivated to continue to be good people despite that fact. They ask: why does a person have to be *capable* of self-perfection?

⁶⁹ This is, of course, acceptable to any ethical thinker who simply thinks that morality does not make such demands on us, and we are misled if we try to make such demands on ourselves. However, for those Kantians (who like Baron) think that morality demands that we continually work to be beneficent and to work towards self-perfection, they would not accept the move of the Self-Improvement view. In this chapter, I aim to charitably work with their intuition (without inconsistency with my view from the previous chapter).

However, morality is importantly different than those other projects. Kant, for good reason, set up the obligation to cultivate our natural talents—musicianship included—in a different way than the obligation to cultivate our moral capacities. In the first case, he said that we ought to cultivate our talents. In the second, he stressed that our duty *is* to ‘be perfect’. We must (on moral grounds) see ourselves as autonomous beings. That self-regard comes prior to discussions of particular duties and capacities. As such, we have to think of ourselves as having the capacity to act as a virtuous person would at any given moment. Since all moments add up to a lifetime, we have to think of ourselves as capable of *being* fully virtuous.

Those who do wrong (even when they are tempted by empirical inclinations) are held responsible because we think of them as having the capacity to do what is right *or* choose to avoid doing what is right (Hill and Zweig 45). When we regard ourselves as rational agents, we think of our wills as autonomous even when we act contrary to the moral law (Herman “CKE” 196). Skeptics about the Kantian account of morality think that this way of thinking about imperfect moral agents is bizarre. They say that we *know* we are imperfectly rational creatures. How can we ever know we have the capacity to choose to do what is right because we are rational agents? Perhaps all that was motivating us all along was something else—perhaps self-interest, perhaps love for others, but quite likely not what ‘reason requires of us’. Moral self-regard is when an agent recognizes herself as “a being with the capacity for self-governance and as a being who has absolute worth because of that autonomy” (Denis *MSR* 171). Yet the skeptics prompt us to question our moral self-regard, because our rational agency may not be up to the task.

We know that we have a number of limitations that compromise our ability to make rational decisions. In my reading of Kant’s move, a belief that we can become fully virtuous is

what assures us of having autonomy. Without that belief, we become suspect to the worries of the skeptics: we may not be rational in the way Kant described (which is necessary for us to be acting autonomously), and perhaps we are fooling ourselves when we regard ourselves as having the capacity to choose to do what is right or not.

An ability merely to improve ourselves alone would allow us to treat ourselves as *autonomish* agents, not as autonomous agents. Yet a lack of commitment to our ability to become perfect would undermine our self-respect. A belief in the possibility of virtue doesn't require us to be unreasonable about the rate of progress: we don't need to assume virtue is going to come easily. Nor do we need to believe in an immortal soul or God (as I will argue). Nonetheless, we must not stop at saying that mere progress is good enough.

3.3 Faith in Humanity View

There is a second option that would direct our moral efforts towards self-perfection: a secular faith in the possibility of moral perfection. In a secular faith, we have a robust attitude of confidence toward some commitment (just not a theological one). Instead of having faith that *I* can become morally perfect over the course of an eternal life, I could have faith that our *community* will reach something like that perfection in the future. Perhaps I can avoid despair by thinking that generations after me will be even better than I am, and eventually will secure something like moral perfection.

Both Andrews Reath and Lara Denis have explored this prospect by mining Kant's discussion of a historical faith in human progress over time. They agree that we need to assure ourselves that we can become perfectly virtuous. Reath and Denis believe that there is some

flexibility in what ‘postulate’ we use. After all, Kant’s arguments do not to show that the postulates (immortality, God) are *true*, but that our belief in them is rational (Surprenant 87).

Kant’s theological version of how we could secure the highest good (virtue and corresponding happiness) was an adaptation of a traditional Christian notion for the immortal soul and a God with all the usual characteristics (Reath 601). On a secular version, virtue and corresponding happiness could come about if we morally educated the next generation better than we were, and they did the same for their children. Over the course of time, they would get closer and closer to moral perfection. It would also require that we (and future generations) work to establish social conditions that support the pursuit of moral activities and the achievement of various moral ends (Reath 615). The happiness of all would be a natural result of our good behavior and just institutions (Reath 615).⁷⁰ We would be “creating just states, forging communities to foster virtue, providing our children with an excellent education, learning more about the world to improve human well-being” (Denis “KCA” 212).

3.4 Response to the Faith in Humanity View

Yet this Faith in Humanity view does not resolve the worry that the Improvement View faced. Recall that we have to think of ourselves as being able to achieve moral perfection in order to make sense of ourselves as autonomous agents—not merely imperfectly rational creatures. If I look to the progress made by the moral community as a whole, rather than by myself as an individual, I am still unable to secure my confidence in my ability to be a fully rational agent. At best, I can regard myself as pursuing *with others* something that *we* may

⁷⁰ Reath is concerned with the role of the highest good in Kant’s moral philosophy: an idea that combines both perfect virtue for all rational beings *and* happiness distributed in proportion to the virtue of those agents. In his paper, he focuses on the *proportionality* clause of the highest good.

achieve (though not in my lifetime) but that *I* will not. This may be fine for my hopes for humanity.

However, the problem at hand is a problem that *I* have: to do my duty. Unfortunately, this historical conception of progress fails to have implications about whether *I* am doing well enough. The result is that I cannot know that my self-respect is merited. I should not take myself to be an autonomous agent, because I may *not* be able to govern myself in the way that virtue requires. Even if I have faith in the moral progress of my community, I am only one link in the chain.

So, the Faith in Humanity looks toward *humanity* being all that it can be. However, their answer is ultimately unsatisfying for each of us, because each of us is answerable to the imperative to “be all that you can be”.

4. My View

Let’s look again at our triad of claims that suggested that ordinary moral thinking contains a contradiction: (1) that we are answerable to the obligation to “be all that you can be”, (2) that it is not possible at any point in time for a person to be all that they can be, and (3) that ought implies can. We have seen two solutions so far: one that rejected claim (2) and the other that rejected claim (1).⁷¹

⁷¹ The other possibility, rejecting claim (3) that “ought implies can”, would be an unconventional strategy in Kantian ethics. I do not have an axe to grind with the idea, nor do I think it’s necessary to reject it for the sake of resolving the trilemma.

My solution is that claims (1) and (2) use different senses of ‘can’. The first uses a ‘can’ of capacity—so in Kantian terms, I have the capacity to be morally perfect, insofar as I am a rational agent. Insofar as we are creatures of the human kind, we have the potential to be the best kind of people. Achieving such an outcome may be out of reach, however! The second uses a ‘can’ of opportunity—I may have the capacity to be morally perfect, but I do not have enough opportunities to become perfect of my kind. If I am imprisoned, or oppressed, or exhausted, I lose all sorts of opportunities to act morally—but I do not lose my capacity.

The apparent contradiction between the three claims rested on an equivocation between the two senses of ‘can’. The first and the third claims use the same sense of “can”—that of capacity. If I do not have the capacity to do something, that means that I should not be held to have an obligation to do it at all. A child lacks many mature capacities, and so we do not hold him to have an obligation to lock up the house at night or care for himself. However, if I do not have the opportunity to do something, that may mean I am *excused*, but it does not mean that I did not have the obligation. I have an obligation to meet a loved one’s distress even when if I am away from them and have no way to be contacted: the obligation remains although I am excused. Obligations can persist despite an agent’s inability to fulfil them.

This is what happens in our puzzle: I have an obligation to “be all that I can be”—no matter how well I am doing, I am always improvable. Yet since “being all that I can be” is not a finite limit, it is not something that at any point in time I can attain. For that reason, we should understand the pull of (1) quite differently. We are answerable to the obligation to “be all that we have the human potential (or capacity) to be”, even though we particular individuals may not be able to realize that potential given that we lack opportunities to do so.

So the puzzle was not the puzzle that Kant, or the Kantians who reacted to him, thought it was. I think that we should reject none of these claims, and I think we can still have a satisfying resolution to what we ought to do. We have a secular way to understand each of the three claims, thanks to the work of the imagination.

All three proffered solutions—Kant’s view, the Self-Improvement view, and the Faith in Humanity view—look beyond what we have now, and toward what we can hope. I think that we should take this insight from those views. In what follows, I will argue that we do best if we reconceive how virtue serves as an ideal for us (here I am inspired by Kant, but not necessarily faithful).⁷² We know that we cannot (as a matter of near certainty) fulfil our obligations of duty in this life. I will ask: what would it take to imagine ourselves as the virtuous person? What work does the imagination do for us? Is that sufficient to avoid despair?

I take for granted that we can come to some understanding of what a virtuous person would be like. We can also imagine a virtuous person in the world: a person who at this moment, and over the course of the rest of their life, acts virtuously. I can also craft a version of that ideal who is like *me*. Her name would be mine; she would have my same physical characteristics and psychological patterns (or at least a close resemblance). I can even partially fill in the story between who I am right now and what the virtuous version of myself would be like. I would have a sense of some attributes and habits of thought I would have to change, even if I don’t know all the ways I would need to change or how I would make those changes. The virtuous me

⁷² I do not feel unfaithful, either—just unwilling to stake that my approach is the view that is most promising to attribute to Kant. However, Kant argues that aiming toward (i) our own virtue and (ii) the happiness of all others are necessary means to counter against my motivations that pull me contrary to what morality require, and that *only* these ends can help me do so (*MM* 6:380-1). Likewise, Kant argues that “it is our universal human duty to *elevate* ourselves to this ideal of moral perfection” (*R* 6:61), yet it is not fully clear what it would take to ‘elevate’ ourselves. What I suggest here might be one good answer.

would do the right thing, right now, out of a broad commitment to doing what the moral law requires. She would do so out of a good and strong will.

The real me is imperfect. I forget, or am distracted, and feel perverse or hangry⁷³ sometimes. Yet I can look to this ideal version of myself when I act, and make-believe that I am relevantly able to act like her. I believe that that imaginative exercise toward a virtuous ideal of ourselves can protect us from despair in two ways: it bolsters us through a lie, and it bolsters us through the truth. First, I will look at the way in which what Tamar Gendler calls “imaginative contagion” may influence our affective states as a result of our imagining in this way. In a sense, this imaginative work comforts us through a lie. Second, I will argue that we get comfort from simply being able to do this imaginative exercise at all.

4.1 The Lie of the Imagination

I acknowledge that this kind of imaginative work towards the ideal is a temporary self-deception—we are not able to believe that perfection is possible for us, but we can make-believe that it is. Nonetheless, I think it is a self-deception that is valuable—it bolsters us to feel secure in our duty and avoid despair.⁷⁴ The imagination is a capacity that is important for us, and it can help us avoid this kind of despair when facing an impossible task.

Imagination is a representative function that can mimic beliefs, desires, or perceptions and sometimes substitutes for them (Currie and Ravenscroft 11). Insofar as imaginings can be similar to these other mental states, behaviors within the imagining tend to conform to the

⁷³ For those unfamiliar with the portmanteau: “hungry” + “angry”.

⁷⁴ Kant thought that self-deception was morally reprehensible. I think that we can use the imagination to, in a sense, deceive ourselves, which would mean that I depart from Kant in thinking that not all forms of self-deception are morally on a par. Thanks to Jordan Mackenzie for her comments here.

patterns of behavior one would perform if one actually had the imagined beliefs, desires, or perceptions (Nichols 7). This is called mirroring. For example, if a child imagines that there is a monster in the closet, she would exhibit all the avoidance-behaviors that she does in cases where was genuinely something fearful there. If she is skilled at vividly imagining the monster, she may well *feel* real apprehension and fear.

Imagination provides us with insights because imaginings have the power to move us in the ways that perceptions, or non-imagined beliefs or desires, can. As illustrated by the case of the monster in the closet, we know that imagining scenarios can activate our affective systems (Nichols 8). The effect is familiar from the social sciences literature: in psychophysiological studies, the technique used to elicit emotions relies on these effects of the imagination. Participants are asked to imagine, or recall, or visualize a situation in which they experienced the target emotion. They are asked to take time to get into the scene, and to do so as vividly as they can (Izard 172). The mental simulation of movement ‘resonates’ in our motor systems, thereby improving motor responses (this is how athletes who ‘visualize’ a target improve their performance) (Decety and Stevens 7, 14).

Vividly imagining that something is true can engage the affective systems even outside the boundaries of the imaginative exercise. Gendler calls this phenomenon “imaginative contagion”. Imaginings can have an influence on our behavior, even apart from the beliefs that we hold to be true in the world. For example, if subjects are asked to put a “cyanide” label on a jar they have just filled with sugar, they will be reluctant to eat any (Rozin and Nemeroff; this example used in Gendler 190). Moreover, there has been an extensive literature on the ways in which that mental practice can improve our performance. Sports psychology has done extensive

research of its effects on motor skills. Mental practice—such as imagining ourselves doing free throws, or improving our golf swing—actually makes a difference to how well we perform.⁷⁵

This effect on real-world motivation and performance is incredibly powerful when we are thinking of a version of ourselves. Hal Hershfield and his fellow researchers ran a research study that asked people to imagine their future selves. The study found that we make better choices now if we are directing our focus toward a version of ourselves in the future. The research was prompted from the observation that emotional responses are heightened when we are given vivid examples: “Donors give more to charity when they hear from a victim; pulmonologists smoke less than other doctors because they see dirty lungs all day” (Hershfield 2013). In the study, software was used to create ‘aged’ versions of participants, who then explored a digital environment where they came face to face with either their current face in the mirror, or their aged face in a mirror. Afterwards, they were asked to allocate \$1,000 among four options: buying something nice for someone special, investing in a retirement fund, planning a fun event, or putting it into a checking account. Those who were exposed to their aged face put twice as much money into the retirement fund than other people (Hershfield 2001 S28).

In another study, Hershfield and his co-researchers, Jean-Louis van Gelder and Loran Nordgren, found that people act more ethically when they feel closer to their future selves. Young adults who’d been asked to write a letter to themselves 20 years in the future were less likely to say they’d make an amoral choice (e.g., buying a stolen laptop) than those who wrote a letter to themselves three months in the future. Likewise, those who were presented with an aged avatar of themselves were less likely than those who saw a current-self avatar to cheat on a test (van Gelder 2013 978).

⁷⁵ See Bird 1984, Creelman 2003 for particular studies of this effect; Frank et. al. gives an excellent literature review.

These results hold true when I vividly imagine myself as a virtuous person as a kind of ‘future self’. We imagine, or in Kant’s words “envisage”, a person who is able to fully abide by the moral law—a moral saint or a sage (*LE* Mrongrovius 29:604-5, *C3* 5:232). A vivid enough make-believe of what it would be like if we were fully rational, capable of making autonomous choices (and choosing to act on the moral law) will serve the appropriate role—without the need to postulate an immortal soul. Given the effects of imaginative contagion, it seems that this kind of practice can give us the motivation to carry on or lift despair. It is a kind of temporary self-deception or pretense.

Kant (and later interpreters like Baron) were right to think through ways in which we can respond to our belief that we should “be all that we can be”. In the Canon of the *First Critique*, Kant tells us that belief involved in his moral religion would be an “attitude for which we regard ourselves as having sufficient grounds, while yet there is no existing means of arriving at certainty in the matter” (*CI* A825/B853). This kind of attitude allowed him to endorse certain religious commitments. By relying on imagination, rather than the faith involved in belief, we can do more than Kant’s moral religion was able to do. Kant’s view was acceptable so long as the evidence was neutral with regard to the existence of God or the immortal soul. But this imaginative directedness towards an ideal of moral perfection who is like me? That can fly in the face of evidence.

We might ask why I should be directed toward myself as a virtuous person, as opposed to other imaginative exercises. I could, for example, vividly imagine myself in circumstances where I have no pressure to act wrongly. All my desires are fulfilled, or people want to coordinate—and I simply have no temptations to do wrong. Yet that will not have the ‘contagious’ effects that imagining myself acting and feeling like the Kantian virtuous agent would in my real life.

Imagination can be used and misused in a number of ways, so let me be clear about what I think it can do for us. The imagination can misfire: I could imagine myself as Wonder Woman because I admire her strength and forthrightness, yet doing so may lead me to behave in ways that are delusional or foolhardy. Yet there are ways to imagine myself that would lead me to do wise things (imagining myself to be a genuinely good person like Mr. Rogers⁷⁶, for example). Certain ideals minimize issues if they have influence in the real world. We don't want to simply make-believe unthinkingly—such behavior could lead to my being insensitive to actual features of the world (my many current imperfections, the needs of others). Nor can I just 'mimic' what my imagined ideal self would do—she is known to be patient, and I am not. People I interact with may respond to the two kinds of people very differently. To use the imagination in this way requires us to act critically.⁷⁷

Some might object to the way that I have detailed the function of the imagination directed towards an ideal of virtue. They may think my response is unsatisfying, because it hangs on self-deception through pretense. However, many valuable activities in our lives depend on self-deception of some kind. Many people rely on a false conception that they are a better partner or parent or person than they actually are. Taken uncritically, this can lead to radical self-

⁷⁶ *Mister Rogers' Neighborhood* was a children's television show that has been a regular part of American public television since 1967. Fred Rogers, the host, used his television show to present young audiences with a message that they had inherent worth and were worthy of love. He was also, by all reports, a remarkable person in his personal life (Merritt).

⁷⁷ Kant warns, for example, that we may go astray when using an ideal "since it can often be defective" (*LE* Mrongrovius 29:604-5). (How exactly it is defective is left to us to sort out—Kant does not say so explicitly. He likely means that misuse of ideals can lead us into an assortment of pathological behaviors: having our 'head in the clouds', thinking that we're doing better than we are and hypocritically judging others, being delusional.) See Thomas E. Hill, Jr., "Is Kantian Ethical Thinking Utopian?" for helpful ways in which we can understand how these kinds of thinking can go wrong, and Chapter 1, Section 4 for my view. Kant thinks that moral agents do best when they try to continue to use and perfect the ideal, though, not to criticize the defects of our working characterization of the morally perfect person (*LE* Collins 27:315-16).

deception.⁷⁸ Yet this practice can also be the right thing to do: it might lead us to avoid worrying about all the ways we do wrong our children or partners; it might influence us to be better, at times. Using the ideal of virtue is immersing in a falsehood, yes, but it is a falsehood that we can recognize. And, if we are careful, it can be a very good one for us to engage in.

4.2 The Truth of the Imagination

While I believe that the ‘contagion’ of imagining ourselves as the ideal of a virtuous person motivates us in an important and meaningful way, I do not think this is the only kind of motivation that the ideal can provide. There is comfort in knowing that something that is ‘impossible’ (out of reach for any person) is still imaginable, even when we are not immersing in an imagining. We can say many things about the virtuous person, and we can imagine ourselves living a life of virtue in a number of ways. That is very different from the virtuous person being a blank to us.

Kant goes this direction as well. He writes, about the ideal, that such a person would be “afflicted by just the same needs and hence also the same sufferings, by just the same natural inclinations and hence also the same temptations to transgression, as we are” (*R* 6:64). If we imagine ourselves as a virtuous person, we would still be much like ourselves—tempted to act in ways motivated by self-interest (even when it would be contrary to what duty requires—but would be the kind of moral agents where we would not end up transgressing).

⁷⁸ In contrast, the ideal of the virtuous person that has been in play in this chapter is assumed to be a person who is reason-governed and so who is mindful of the dangers of uncaring idealism. Such a person makes adjustments to live well in a non-ideal world. If we are to imagine ourselves as perfectly virtuous, we would also perhaps do a better job in deciding what is right to do. We might be able to think better, see the options more clearly. As such, the imagination would serve practical reason. Thanks to Tom Hill for this point.

That kind of imaginability is comforting. As Iris Murdoch says, “The human mind is naturally and largely given to fantasy” (Murdoch 322). This propensity to fantasy has bad results, such as vanity and delusions of grandeur (so again, we must be careful in how we exercise our imaginations). But there would be a special kind of hopelessness when we know that we cannot even imagine what something so valuable would even be like.

Empirical grounds are sufficient to convince us that being our best is not ‘possible’ in the sense of being difficult to achieve but in reach. It just seems like something that could never be achieved. As a result, we cannot ‘hope’ for moral perfection in the way that we can hope for things as an extension of the present. I can hope that my day will go well or even that my life-plan will succeed. In such cases, our hopes are turned toward “a belief that we can, through our own actions, make something turn out *better* than would otherwise be expected” (Reading 4).

Some of our hopes are directed towards a future that is not connected to our present—outcomes where there is no clear path between now and the hoped-for circumstance. Some might even consider this the more important sense of hope. St Paul, for example, wrote “Hope that is seen is not hope: for what a man seeth, why doth he yet hope for?” (Romans 8:24; reference from Reading 6). This is the kind of hope engaged when we imagine the ideal of virtue—in which we may not be able to securely see how what we hope for could come about.

If we can describe the conditions under which the ideal of virtue would come about, however, it can secure hope for our moral life and embracing the moral life. There is comfort in knowing that the virtuous person is possible in the sense that while it is incredibly remote, it is accessible to the imagination.

Murdoch uses this technique to respond to those intellectuals who anticipate a terrible future for us, and despair about it (Murdoch 208). She thinks that science fiction, for example, is

art that responds to this despair. “Science fiction [...] soothes our disturbed frustrated imagination by picturing our more distant surroundings as being just like our domestic ones only, in our local terms, more weird; and though it may strain every nerve it cannot do otherwise.” (Murdoch 208). Science fiction leads us to feel less pessimistic about what the future holds, *because we can imagine a future in which our lives are much like they are now*. That is far better than not being able to predict what the future might be like at all, or only being able to picture grim futures.

The imagination is deeply creative. “Imagination concerns the depth and working of the human soul and its truthful visions. [...] Human life mingles horror and absurdity, and great art has the ability to mingle these and do justice to them in the process.” (Murdoch 208-9). I suspect that everyday activities, such as thinking about ourselves as an ‘ideal’, might be able to do much the same as art. We know that we can be wretched beings. Yet we may be able to do some justice to ourselves when we imagine ourselves as the ideal.

Conclusion

I have examined our set of three apparently inconsistent claims, and suggested that rather than rejecting one of them (as proposed by Kant as well as various Kantians), the puzzle dissolves if we take a closer look at the kind of ‘can’ involved in claims (1) and (2). The first claim—that we ought to “be all that we can be”—used a ‘can’ about what capacities we have. The second claim—that it is not possible for a person to be all that they can be at any time—used a ‘can’ of opportunity. Moreover, by exploring how our imagination encourages us to act in ways that are sometimes contrary to evidence, I have offered a way to understand our attraction

to each of the claims—and how we can satisfyingly think of ourselves as being motivated to act on (1) while recognizing the truth of (2). The work done by the imagination is a resource for both theorists and non-theorists who take the slogan “be all you can be” seriously.

WORKS CITED

- Anderson, M. T. *Symphony for the City of the Dead*. Somerville, MA: Candlewick Press, 2015.
- Baron, Marcia. "Kantian Ethics and the Supererogatory" ["KES"] in *Kantian Ethics, Almost Without Apology*. Ithaca: Cornell University Press, 1995, pp. 21-58.
- Baron, Marcia. "The Supererogatory and Kant's Imperfect Duties" ["SKID"] in *Reason, Value, and Respect*, edited by Mark Timmons and Robert N. Johnson. New York: Oxford University Press, 2015.
- The Bible, King James Version
- Bird, Evelyn I. "EMG Quantification of Mental Rehearsal" in *Perceptual and Motor Skills*, Vol. 59 (1984), pp. 899-906.
- Biss, Mavis. "Kantian Moral Striving" in *Kantian Review*, Vol. 20, No. 1 (2015), pp. 1-23.
- Creelman, Jim. "Influence of Mental Practice on Development of Voluntary Control of a Novel Motor Acquisition Task" in *Perceptual and Motor Skills*, Vol. 97 (2003), pp. 319-337.
- Currie, Gregory and Ian Ravenscroft. *Recreative Minds: imagination in philosophy and psychology*. New York: Oxford University Press, 2002.
- Decety and Stevens, "Action Representation and Its Role in Social Interaction", Chapter 1 of *The Handbook of Imagination and Mental Simulation*, edited by Keith D. Markman, William M. P. Klein, and Julie A. Suhr. New York: Psychology Press, 2009.
- Denis, Lara. *Moral Self-Regard*. New York: Garland Publishers, 2001.
- Denis, Lara. "Kant's Criticism of Atheism" in *Kant-Studien*, Vol. 94, No. 2 (May 2003), pp. 198-219.
- Frank, Cornelia et. al. "Mental Practice and Mental Rehearsal: Experimental Investigation on the Functional Links between Motor Memory and Motor Imagery" in *Plos One*, Vol. 9, Issue 4 (2014), e95175.
- Gendler, Tamar. "Imaginative Contagion" in *Metaphilosophy*, Vol. 37, No. 2 (April 2006), pp. 183-203.
- Hedley, Douglas. *Living Forms of the Imagination*. New York: T&T Clark, 2008.
- Herman, Barbara. "A Cosmopolitan Kingdom of Ends" in *Reclaiming the History of Ethics: Essays for John Rawls*, edited by Andrews Reath, Barbara Herman, and Christine M. Korsgaard. New York: Cambridge University Press, 1997, pp. 187-213.

- Hershfield, Hal, interview by Alison Beard. “You Make Better Decisions if You “See” Your Senior Self” in *The Harvard Business Review* (June 2013). Online: <https://hbr.org/2013/06/you-make-better-decisions-if-you-see-your-senior-self>
- Hershfield, Hal E. et. al. “Increasing Saving Behavior Through Age-Progressed Renderings of the Future Self” in *Journal of Marketing Research*, Vol. XLVIII (Special Issue 2011), pp. S23-S37.
- Herman, Barbara. *The Practice of Moral Judgment*. Cambridge, MA: Harvard University Press, 1993.
- Hill, Jr., Thomas E. “Is Kantian Ethical Thinking Utopian?” (manuscript)
- Hill Jr., Thomas E. and Arnulf Zweig, “Editors’ Introduction” to *Groundwork for the Metaphysics of Morals* by Immanuel Kant, edited by Thomas E. Hill, Jr. and Arnulf Zweig. New York: Oxford University Press, 2002.
- Izard, Carroll E. *Psychology of Emotions*. New York: Plenum Press, 1991.
- Kant, Immanuel. *Critique of Pure Reason [C1]*, translated by Norman Kemp Smith. London: Palgrave Macmillan, 2007.
- Kant, Immanuel. *Critique of Practical Reason [C2]*, translated by Lewis White Beck. New York: Macmillan Publishing Company, 1956.
- Kant, Immanuel. *Critique of the Power of Judgment [C3]*, edited by Paul Guyer, translated by Paul Guyer and Eric Matthews. New York: Cambridge University Press, 2000.
- Kant, Immanuel. *Groundwork for the Metaphysics of Morals [G]*, translated by Arnulf Zweig, edited by Thomas E. Hill Jr. and Arnulf Zweig. New York: Oxford University Press, 2002.
- Kant, Immanuel. *Lectures on Ethics [LE]*, edited by Peter Heath and J.B. Schneewind, translated by Peter Heath. New York: Cambridge University Press, 2001.
- Kant, Immanuel. *Metaphysics of Morals [MM]*, translated and edited by Mary Gregor. New York: Cambridge University Press, 1996.
- Kant, Immanuel. *Religion within the Boundaries of Mere Reason [R]*, edited by Allen Wood and George di Giovanni. New York: Cambridge University Press, 1998.
- Mariña, Jacqueline. “Making Sense of Kant’s Highest Good”, *Kant-Studien*, Vol. 91 (2000), pp. 329-355.
- Merritt, Jonathan. “Saint Fred” in *The Atlantic*. November 22, 2015. Online: <https://www.theatlantic.com/politics/archive/2015/11/mister-rogers-saint/416838/>
- Murdoch, Iris. *Metaphysics as a Guide to Morals*. London: Penguin, 1993.

- Nichols, Shaun. "Introduction" to *The Architecture of the Imagination*, edited by Shaun Nichols. New York: Oxford University Press, 2006.
- Reading, Anthony. *Hope and Despair*. Baltimore: The Johns Hopkins University Press, 2004.
- Reath, Andrews. "Two Conceptions of the Highest Good in Kant", *Journal of the History of Philosophy*, Vol. 26, No. 4 (October 1988), pp. 593-619.
- Rozin and Nemeroff. "The Laws of Sympathetic Magic: A Psychological Analysis of Similarity and Contagion" in *Cultural Psychology: Essays on Comparative Human Development*, edited by J. Stigler, G. Herdt, and R. A. Shweder, pp. 203–32. Cambridge: Cambridge University Press, 1991.
- Surprenant, Chris W. "Kant's Postulate of the Immortality of the Soul" in the *International Philosophical Quarterly*, Vol. 48, No. 1, Issue 189 (March 2008), pp. 85-98.
- Van Gelder, Jean Louis, Hal E. Hershfield, and Loran F. Nordgren. "Vividness of the Future Self Predicts Delinquency" in *Psychological Science*, Vol. 24, No. 6 (2013), pp. 974-980.
- Wood, Allen. "Kant's history of ethics" in *Kant's Lectures on ethics: A Critical Guide*, edited by Lara Denis and Oliver Sensen. New York: Cambridge University Press, 2015, pp. 126-34.

CHAPTER 4: THE KINGDOM OF ENDS AS AN IDEAL

The kingdom of ends is known from Kant's *Groundwork to the Metaphysics of Morals* as a "systematic union" of rational agents, each of whom is engaged in setting laws for the community and must abide by those laws (*G* 4:433, 4:462). It also is one of the few concepts that Kant describes as an 'ideal' in the *Groundwork*.⁷⁹ As such, we might take the kingdom of ends to be an appealing utopian vision of what such our community could look like. Moreover, the ideal of the kingdom of ends *as a community* may help Kantian ethics resist the criticisms that its emphasis on autonomy leaves no room for us to appropriately think of ourselves as social beings.⁸⁰

The kingdom of ends is also of interest to interpreters insofar as it is the center of the Formulation of Kingdom of Ends (FKE), the last variation of the Categorical Imperative, which reads: "All maxims which stem from autonomous lawgiving are to harmonize with a possible kingdom of ends" (*G* 4:436). This formulation of the Categorical Imperative illustrates what it is to *give law*, or to *legislate*, for a community of equal moral agents that includes oneself. If the kingdom of ends serves as an ideal for deliberation and legislation, then the kingdom of ends might be of great use for both ethics and for political philosophy.

⁷⁹ The only other ideals besides the kingdom of ends mentioned in the *Groundwork* are those of moral perfection (4:408) and happiness (which Kant says is an ideal of the *imagination*, not of reason) (4:418). Elsewhere in the ethical writings Kant also describes the highest good (*R* 6:135) and friendship (*MM* 6:469) as ideals. Finally, in the *Critique of Judgment*, he also includes beauty as an ideal (though not an ethical ideal) (*C3* 5:229ff).

⁸⁰ See Herman, "A Cosmopolitan Kingdom of Ends" and Reath, "Legislating for a Realm of Ends: The Social Dimension of Autonomy" for versions of this position.

The language of the ‘kingdom of ends’ is not clearly carried over into Kant’s later ethical writings. Nonetheless, it is reasonable to think that the kingdom of ends refers to the same ideal as the ‘ethical community’ described in the *Religion* (*R* 6:96, 6:101),⁸¹ which focuses on the ways in which moral agents could act in community with each other.⁸² I will take this continuity between the two concepts for granted, and use passages from both the *Groundwork* and the *Religion* to elucidate our topic.

While Kant clearly specifies that he takes the kingdom of ends to be an ideal (*G* 4:433), he offers little further detail of *in what sense* it should serve as an ideal. After setting out Kant’s description of the ethical community (Section 1), I will distinguish two interpretations for kingdom of ends: as a *substantive* ideal, which serves as a standard of excellence that often can help guide improvement of a given target (in this case, a community of persons), and as a *deliberative* ideal, which specifies what conditions must hold for deliberating on and deciding ethical questions. In Section 2, I will explore how Kant interpreters have focused on either one or the other option. The remainder of the paper will discuss the ways in which the kingdom of ends could be interpreted *as an ideal* and what benefit that might have for Kantian ethics.

⁸¹ The ideal of the ethical community is referred to *as* an ideal at *R* 6:96 and 6:101. The other ideals referenced in the *Religion* include that of moral perfection or holiness (*R* 6:61, 64-7, 76, 129, 191) and the highest good (*R* 6:135). There are so few concepts identified *as* ideals across Kant’s ethical writings that it is likely that the kingdom of ends (a kind of ethical community) of the *Groundwork* and the ethical community of the *Religion* are referring to the same ideal.

⁸² Part III of the *Religion* is devoted to talking about the ways in which we engage with one another in community. While Kant is, in this text, dedicated to exploring moral religion (and speaks of the ethical community in terms of a ‘church’), there are a number of ideas expressed that fall in line with the original commitments of the *Groundwork*’s kingdom of ends. The language of a “kingdom” is carried over into the *Religion* to describe this community—although as a “Kingdom of God” rather than a “kingdom of ends” (*R* 6:151-2). Allen Wood, for one, seems to read the *kingdom of ends* and the *ethical community* (as well as the *highest good*) as one and the same. Each of these are described as the “systematic union of rational beings under common objective laws” (Wood 188, referring to *R* 6:98, *G* 4:433). Like in the *Groundwork*, the ethical community of the *Religion* is described as a matter of public legislative action, under a common lawgiver (*R* 6:98-99, *G* 4:433).

1. Kant on the Ethical Community

1.1 *The kingdom of ends: as a community and as a legislative strategy*

The ‘kingdom of ends’⁸³ is a difficult concept to pin down. It is introduced to us in the *Groundwork* as a “systematic union of different rational beings under common laws” (*G* 4:433). The second part of the description is easier to grapple with than the first. A union of persons under common laws is exactly what we expect when we describe a ‘kingdom’—a *community of persons* of some kind.

The ethical community seems like a desirable focus for our attentions. Up to this point of the *Groundwork*, Kant explored how we should test our maxims to see if they could function as universal laws (*G* 4:421), and that we should respect others as ends in themselves (*G* 4:428). However, those foci for how to understand the moral law do not give us exact guidance for what it would be like for rational agents to collectively act according to the moral law. We might worry, for example, that agents acting on the moral law might conflict with each other in executing their duties. Or we might worry that we don’t know what relationships between persons look like when all are acting on the moral law.

The ethical community that Kant describes is one of ideal citizens who both serve as ideal *legislators* of laws for all persons in the community but are also *subject* to those laws (*G* 4:433; Hill and Zweig 87). These ideal citizens look very unlike actual persons and our actual communities. Unlike actual persons, these citizens would act in ways that are perfectly rational. They would set for themselves and obey the moral law (*G* 4:435). Perhaps, then, it is best to conceive of these ideal persons as actual persons “as they could be” (Hill and Zweig 87). Unlike

⁸³ I have already staked an interpretive claim by reading the ‘kingdom of ends’ from the *Groundwork* as the same concept as the ‘ethical community’ from the *Religion*. For my reasons to do so, see footnotes 3 and 4.

our local, national, or global communities, each person (by virtue of their status as a rational agent) serves as a legislator instead of, like the for most of us, as a subject only.

These persons stand in moral relationships to each other. They are able to live under laws that respect their dignity as ends in themselves, as well as promote the (permissible) personal ends that each sets for themselves (*G* 4:433). As Robert Adams describes the ethical community of the *Religion*, the purpose of this community is to provide a social structure “in which people instruct, encourage, and support each other in virtue, instead of providing each other with temptations to vice” (Adams xxviii). Given the ways in which the members of the ethical community are subject to the same law—the moral law—they are well-coordinated (they have a ‘unity of purpose’, in the language of the *Religion*) (6:96).⁸⁴

Nonetheless, the first part of Kant’s description—a “systematic union” of these rational agents—needs some further attention. What seems to provide this systematicity is the ways in which we *legislate* together as rational agents.⁸⁵ Kant describes ‘legislation’ as choosing our maxims from the point of view of both oneself and every other rational being (*G* 4:438). So in legislating, we acknowledge the dignity of each rational being as a law-maker themselves (*G* 4:434). We must legislate without consulting what would be in our self-interest (*G* 4:432). If we

⁸⁴ One preliminary question we might want to ask is whether the kingdom of ends includes everyone, virtuous or not, or only the virtuous? I choose to follow Hill and Zweig’s suggestion that there would be an actual kingdom of ends *only if* everyone were to follow the moral law—so the kingdom of ends includes everyone, but everyone conceived of as virtuous (Hill and Zweig 88, referring to *G* 4:438). Similarly, Allen Wood writes that in a ‘minimal sense’, we are already an ethical community. However, there is a fuller sense in which we could become an ethical community by *realizing*, in practice, a moral unity with all other rational agents (Wood 191). Herman, in “A Cosmopolitan Kingdom of Ends”, gives a very different reading: she thinks that there is a difference between (a) the autonomous legislating members of a kingdom of ends, (b) a kingdom of *good* wills, and (c) the highest good (good wills as well as happiness) (Herman 64).

⁸⁵ It should not be forgotten that the discussion of the kingdom of ends immediately follows, and is presented as a kind of variation of, the Formula of Autonomy. The Formula of Autonomy describes us as each having a will that “legislates universal law” (*G* 4:431). This is one important reason to think the legislative features of the kingdom of ends are important!

are to legislate from the point of view of a community of persons, we must be considerate and thoughtful in certain ways.⁸⁶

One important point is that this discussion of *legislation* might lead us to think of the kingdom of ends as a kind of *political community*. However, we must be careful about Kant's meaning here. A political state enforces laws of justice, by limiting each person's freedom so that it can be consistent with the freedom of all others. It uses various kinds of sanctions or penalties to do so (Wood 190, referring to *R* 6:98). The ethical community, in contrast, helps to teach *voluntary* compliance with laws of virtue independent of any kind of political authority (*R* 6:95). Laws of virtue cannot be motivated or enforced by any political actor or institution, but must be chosen and acted upon by each person (Adams xxviii, Wood 189).⁸⁷ 'Legislation', then, must be understood in a somewhat different way than in our ordinary political institutions.

1.2 The kingdom of ends as an ideal

The kingdom of ends is an ideal in Kant's ethical theory (*G* 4:462, *R* 6:96). An *ideal*, in Kant's technical sense, is related in a special way to an *idea*.⁸⁸ In the *Lectures on Ethics*, ideas are:

“a moral perfection whose object can never be adequately given in experience. [...] they constitute the guideline to which we must constantly approach. They make up the law of approximation. We have to possess a yardstick by which to estimate our moral worth, and

⁸⁶ Kant's discussion of 'legislation' is notably minimal in this passage. For this reason, interpreters (such as Tom Hill) have given much more substantial accounts of what it is to 'legislate' (see Section 2.2).

⁸⁷ The kingdom of ends can also be helpful in political philosophy—both John Rawls and Christine Korsgaard have used it for those purposes (Korsgaard “I” xxviii). Nonetheless, I think it is important to *not* reduce the ethical community to a kind of democratic republic.

⁸⁸ The technical notion of the Ideal was first stated in the *Critique of Pure Reason*. The *Critique of Pure Reason*'s description of the distinction between ideas and ideals is, unfortunately, opaque (*C1* A567-71/B596-99). I have instead chosen to use the more accessible statements from the third *Critique* and the *Lectures on Ethics*. See Richard Dean's “Humanity as an Idea, as an Ideal, and as an End in Itself” for an exegesis of Kant's distinction between an idea and an ideal.

to know the degree to which we are faulty and deficient; and here I have to conceive of a maximum, so that I know how far away I am, or how near I come to it.” (*LE* Mrongrovius 29:604-5).

In short, we can understand ideals as concepts. For example, we have an *idea* of moral perfection (as discussed in Chapter 2): for a rational agent to have a good and strong will to follow the moral law *because* it is the moral law. That is a concept of virtue—it does not connect with empirical features of the world, according to Kant.

Each ideal corresponds to some idea. Kant writes, “All ideals are fictions. We attempt, *in concreto*, to envisage a being that is congruent with the Idea. In the ideal we turn the Ideas into a model, and may go astray in clinging to an ideal, since it can often be defective” (*LE* Mrongrovius 29:605).⁸⁹ Thus, the ideal is an instantiation of the idea that has empirical qualities. So continuing our earlier example, we can apply the *idea* of moral perfection to a particular object: a person. The *ideal* of the virtuous person is an individual person who possesses everything that we would need to be good: “goodness of soul, or purity, or fortitude, or serenity, etc.” (*C3* 5:235). An ideal can serve as a model of a *realistic* person.⁹⁰ As Kant writes, “As the idea gives the *rule*, so the ideal in such a case serves as the *archetype*” (*C1* A569/B597).⁹¹

Ideals are special tools in our moral life. They require a combination of ideas of reason and “great force of imagination” in both our judgment of them or our presentation of them (*C3* 5:235). They can help us demonstrate to ourselves that what we *ought* to do is possible for us (*R* 6:62). The fact that one can imagine a human being who can conduct themselves in a morally

⁸⁹ The *Lectures on Ethics*’ description is more user-friendly, but it should be noted the source isn’t as trustworthy, however, because it is student notes, not published work by Kant. Nonetheless, it seems to correspond to the idea/ideal description in the first *Critique* (*C1* A567-8/B595-6ff) and the later third *Critique* (*C3* 5:232).

⁹⁰ Remembering, always, that we should be wary about the ways in which ideals can lead us astray. For a general discussion of this topic, see Chapter 1, Section 4.

⁹¹ In some ways, this distinction between idea and ideal might remind us of the Chapter 1 distinction between a stipulative and imaginative ideal.

perfect way—Kant gives a version of what is recognizably a Christ figure for this ideal—can help “attune our mind” as a proof that such a moral goodness “can be practiced and attained by us” (*R* 6:64).

Similarly, we might ask how the ideal of the kingdom of ends can serve us as a model. First, what is the corresponding *idea* of reason that is applied in the ideal of the kingdom of ends? Our most obvious candidate is the idea that Kant articulates immediately prior to the introduction of the Formula of the Kingdom of Ends: the idea of *every* person as a legislating agent (*G* 4:431-2). As an *ideal*, then, the kingdom of ends doesn’t describe or refer to any historical community nor one that we can expect to see later in our world (Hill and Zweig 86). Instead, it is a model of a realistic set of agents⁹² who are able to unify their moral action via legislation. So the kingdom of ends *interests* us because it serves as a social ideal: not just for an individual agent (as the ideal of the virtuous person does), but for a world of rational agents.

Attempting to bring about the ideal ethical community is a duty for us—in Kant’s words, “to unite in an ethical community is a duty of a special kind” (*R* 6:151). This is a duty “to work toward a moral community of its members, to take responsibility for all men as persons, and for the moral relations *between* men” (Wood 190). So we might want to ask ourselves how the kingdom of ends generates particular duties. I will survey three⁹³ options, and in the remainder of the chapter draw out two of these in detail.

⁹² Realistic in the sense that these virtuous agents are conceived of as being subject to temptations—they have emotions and empirical needs that they must satisfy—but unlike us, they would be virtuous (*R* 6:64-5). This is very different than picturing a community of rational agents without those susceptibilities!

⁹³ A fourth way, which I will not discuss here, is to aim for a harmony of human and natural purposes (an option mentioned in Hill and Zweig 89, referring to Paton’s strategy from *The Categorical Imperative*, pp. 185-98).

One way to understand the kingdom of ends is that it gives us an imperative to approximate an ideal state of affairs: to work for a community of a particular kind (Wood 189).⁹⁴ In the *Religion*, Kant writes that “the dominion of the good principle is not otherwise attainable, so far as human beings can work towards it, than through the setting up and the diffusion of a society in accordance with, and for the sake of, the laws of virtue—a society which reason makes it a task and a duty of the entire human race to establish in its full scope” (*R* 6:94). So perhaps we should set up our ordinary communities in a way that more and more resembles the ideal ethical community.

A second way is to think of the kingdom of ends as deliberative procedure: “an abstract representation of an appropriate standpoint for deliberating with others about general moral principles—what they require in general, what exceptions they should include, and why” (Hill and Zweig 89). An ideal deliberator would be fully rational, would respect humanity as an end in itself, would set aside their own special interests, would count laws as valid only when justifiable to all, and would accept that we are all subject to the laws (Hill and Zweig 89).⁹⁵ As such, this might move us away from what *I* can will as universal law, but what any rational or reasonable person would will (Hill and Zweig 89).

One final option is to simply say that we should simply do our duty from duty (as instructed by the earlier formulas of the Categorical Imperative), and to think that the kingdom of

⁹⁴ Wood notes how we might need to have faith in God in order to actually *secure* the highest good (‘proportionate happiness’ is very difficult for us to secure), but nonetheless we shouldn’t simply just sit back and wait for God to do such work for us (Wood 189).

⁹⁵ Hill and Zweig note that there are remaining issues to solve, given this brief description. In particular, we know that there is deep disagreement on moral principles between persons. How do we deal with these real-world disagreements? Second, how do we generate principles for action? The principles that would be appropriate for a world of virtuous persons are likely to be inappropriate or disastrous in our actual world (Hill and Zweig 90).

ends does not generate any new requirements (Hill and Zweig 89).⁹⁶ I will set this option aside for present purposes. It seems most natural to take Kant's instructions around the kingdom of ends—that we *must* regard ourselves as legislators in a kingdom of ends—at face value (*G* 4:434).

2. Interpretations of the Kingdom of Ends

We thus have a few options as to how to interpret the kingdom of ends as an ideal. Two of those options hook on to descriptions in Kant's ethical writings where he focuses on how the kingdom of ends serves as an ideal of a *community* of agents, or alternately, how it could serve as guidance for how we ought to *deliberate*. In this section, I will focus in on how these two interpretations have been offered: (a) the kingdom of ends as a *substantive ideal* for what it is to be an ideal moral community, and (b) the kingdom of ends as a *deliberative ideal* for individual persons living within the community. On the first interpretation, one would read the kingdom of ends as guidance for how we should set up and conduct ourselves as a community; on the second, one would read the kingdom of ends as articulating the proper way that a moral agent should deliberate about what they should do, given that they are one person in a larger moral

⁹⁶ Some interpreters, including Onora O'Neill, read the series of formulations of the Categorical Imperative such that all of the later formulations ultimately reduce to the Formulation of Universal Law (FUL). If those interpreters are right, then we should not take the Formulation of the Kingdom of Ends (FKE) as generating any further duties. While substantial discussion of the textual grounds for any interpretation of the relationship between the formulas is beyond the scope of this chapter, even O'Neill argues that the notion of a plurality of interacting agents is part of what we are considering, even in the Formula of Universal Law (O'Neill 94). That is the minimum needed for this chapter.

community.⁹⁷ Let us take each of these in turn, and see why they might be appropriate ways of conceiving of how the kingdom of ends could serve as an ideal.

2.1: A substantive ideal

A majority of interpreters seem to take the kingdom of agents as, primarily, an ideal community of agents. These include Barbara Herman, Christine Korsgaard, Kate Moran, Andrews Reath, and Allen Wood.⁹⁸ We know at least some of the reasons why it is evident in the text that the kingdom of ends should be taken as a substantive ideal⁹⁹ for what it is to be an ideal moral community—that is, to “form an *idea* of a moral community of men as a goal of our actual social endeavor” (Wood 191-2). The corresponding ideal helps us to recognize *ourselves* as living with others: “to think of finite beings, in a place, with possessions, attachments, histories, and the rest.” (Herman “CKE” 202).

There are two major reasons why we might think of the kingdom of ends as an ideal of what a community should strive to be like. *First*, we know that we have interests and projects that are in conflict with one another. Both moral and political philosophers are interested in how,

⁹⁷ Yet another reading, which I will not examine in this chapter, can be found in the work of Katrin Flikschuh, which gives what she calls a “metaphysical” reading of the ideal. She thinks that the kingdom of ends is not a “blueprint for an empirically realizable moral order” (Flikschuh 137). Instead, the kingdom of ends is analogous to the ideal of the highest good: “which similarly functions to sustain individuals’ practical faith in morality in the face of sensibly conditioned adversity” (Flikschuh 137). She thinks that the formulation of the kingdom of ends lacks the element of coordination, unlike what I have set here, and so is committed to thinking that the kingdom of ends cannot aid us in directing ourselves to a harmonious moral community (Flikschuh 137).

⁹⁸ See Korsgaard, *CKE* p. 192; Moran, p. 76; Reath, “LKE” p. 173; and Wood, p. 191-2.

⁹⁹ The language “substantive” ideal may not be perfect to capture the distinction here. Nonetheless, it is intended to capture something about a reasoning being (or set of reasoning beings) *as* an object, as contrasted to thinking about a reasoning being (or set of reasoning beings) *as* a reasoner (or reasoners). We might be able to elicit the same distinction if we compare how philosophers discuss the virtuous person using questions like “What would they be like? What virtues would they possess? What could we observe about them?” as opposed to using questions like, “How would they reason about a moral situation? What kinds of facts would be salient to them? How might their emotions move them?” Both sets of questions may be important, but one is certainly more ‘from the inside’.

exactly, we are to resolve these tensions. *Second*, as Herman points out, the kingdom of ends can show how we both can prioritize autonomy *as well as* deep social connections between Kantian agents. This is because the Kantian rational agent is not an isolated individual, but a rational agent living together with others (Herman “CKE” 188).

2.1.1 Trust and Mistrust of Others

The kingdom of ends is a community of agents who are *rational* but have *empirical needs*. Our rational capacities allow us to serve as legislators in the kingdom of ends. However, we are also *empirical* beings—we have needs and inclinations, and must ensure that we meet those needs and inclinations in ways that are consistent with what the moral law requires of us (Herman “CKE” 196; *G* 4:433-4). Perhaps one way to think of the kingdom of ends as a ‘systematic unity’ is that it both articulates our commitment to treat other rational agents *as* fellow law-makers, but also offers us a way to harmonize those needs (consistent with treating other agents according to their status).

We have many empirical needs that make us dependent on certain things to be true in the world: needs for food and drink, susceptibility to injury and death. We need conditions in which we can secure sustenance, medical care, shelter, and so on—and that is just the bare minimum for us to continue to live. In order to secure this minimum of goods required to meet our needs, we require (at least) non-interference from others. In many cases, we need *assistance* from others in order to secure these goods (Herman “CKE” 197). We regularly need others’ assistance in order to grow or secure food, craft shelters, and provide medical care.

Moreover, we are *interested* in being with others. It is an empirical fact that we need others in order to secure minimal conditions for our own happiness and that we often find mutual advantage in cooperating with others. Yet we also are social beings who want to *be* with other

persons (*MM* 6:471). We want to be able to share and discuss our thoughts with others. We want to love and engage with others (*MM* 6:469). We want sympathy from others (*G* 4:423). All of these social interests require, by definition, being with others.

Unfortunately, we are terrible at playing nice with others. We mistrust other people—and often for good reason! Each of us is what Kant calls ‘frail’ (that is, we are liable to fail to have the strength to comply with the principles we have adopted) and we are dishonest in scrutinizing our own motives (*R* 6:37). We are regularly tempted to take advantage of others (*G* 4:424). Given that our motives are opaque even to ourselves, we might not even *realize* that we are aiming to take advantage of others. We are fearful of what others might do or say about us if they were to know too much about our secrets or our private judgments (*MM* 6:471). Even in the closest friendships, we find it to be “prudent” to have some limit on trust, even between the closest friends (*R* 6:34). We also have a tendency to devolve into vicious attitudes: we can hate the people who have helped us and feel *schadenfreude* even with regard to the misfortunes of our closest friends (*R* 6:34).

These features of what we are like might give us a *prima facie* reason to read the kingdom of ends as a substantive ideal for a community. We need some kind of guidance for how to do well while living with others. We know what it would be like for us to become more virtuous—to secure both a good and strong will, and to be better in our self-scrutiny. However, we may well wonder what our interpersonal engagements might look like, were we to be virtuous. What might trusting or loving relationships look like between virtuous persons? How might those virtuous persons arrange their community affairs?

The kingdom of ends as a *substantive* ideal for a community might be able to articulate what it would be like *if only* each of us were to act as the moral law required. We could imagine

ourselves without such temptations to take advantage of others, because not only would we have a *good* will, we would have a *strong* good will (that is, a virtuous one). We could imagine ourselves without fear that others may harm us—that we could *securely* trust other persons who are virtuous. Imagining the kingdom of ends bolsters us to avoid despair that the flaws of our ordinary interpersonal interactions wouldn’t remain amongst a community of virtuous agents. As I will discuss in Section 5, there are ways in which imagining so can prompt realistic thought exercises about what actions it would take to make our world more like the kingdom of ends, and would also dispose us to act as one would need to do so in order to bring about such an ideal moral community.¹⁰⁰

2.1.2 Features of the substantive ideal

The interpreters who present the kingdom of ends as a substantive ideal for a moral community point to these two features: how it could help us (1) avoid conflicts and secure coordination, and (2) secure our social connections with others.

(1) Again, an empirical fact about humanity is that we each want to secure our own happiness. We know that we are motivated to act for many and varied ends—some compatible with doing our duty, others not. One way we can make these projects and interests mutually consistent across persons is by accepting ‘side-constraints’ of permissibility. So, for example, I won’t carry out my project if it would injure you (or otherwise breach a perfect duty that I have to other persons) (Herman “CKE” 207-8). I must rein in my self-interest because of its tendency to tempt me to undermine others’ rational agency.

¹⁰⁰ In particular, the by-now familiar features of the imagination (mirroring and contagion) from Chapter 1 can be understood to do this kind of work. See in particular Sections 5.1 and 5.2 for a discussion of the features of and constraints on the imagination.

I must also satisfy whatever positive moral obligations apply. For example, I can carry out my projects, so long as I am also working to help others secure their own happiness and to promote my own perfection (Herman “CKE” 207-8). For example, “in circumstances of scarcity, when my action would have an impact on vital resources, I may not consider it as “merely an economic venture.”” (Herman “CKE” 204). We must, then, believe that some kind of cooperation is possible with others. Imagining a substantive ideal of a community of human agents might give us some kinds of instruction for how to secure cooperation and avoid conflict.

(2) However, we are not only beings with these basic needs. If that were the case, we would merely need some kinds of rules and institutions that would resolve conflicts and permit cooperation (*G* 4:405). Another fact about ourselves is that we have deep social connections with others.¹⁰¹ Herman speaks to how it is important for individual rational agents to have other rational agents to be around (Herman “CKE” 188). This kingdom of ends “plainly gives expression to some social element in Kant’s account of moral judgment and obligation” (Herman “CKE” 190).

An ideal of a social order is helpful, here: it could demonstrate how we would act with others in a way that corresponds to the moral law. The ideal of the kingdom of ends tells us that even though we as persons vary immensely about what we care about and set as private ends, we still can enter a moral order (Herman “CKE” 207) We can consider how respecting each person

¹⁰¹ Herman thinks that Kantian ethics is attractive because of “the dignity it accords the individual person because of her capacity to act both freely and as reason requires, as well as from the moral equality of persons that follows from locating the basis of moral status in each agent’s own practical reason” (Herman “CKE” 187). Yet it looks to some objectors to be, by this theoretical emphasis, be supportive of a kind of individualism, with persons who are radically separate from one another (Herman “CKE” 188). They worry that we cannot have both autonomy and deep social connection within Kantian ethics (Herman “CKE” 188). Herman argues against these objectors in “A Cosmopolitan Kingdom of Ends”.

as an end in themselves would or might work out in conditions like ours to form particular communities, relationships, or sets of specific rights and duties.

Interpreters have found the reading of the kingdom of ends as a substantive ideal to be compelling.¹⁰² This reading would allow us to imagine what it would be like if “Each citizen takes his own perfection and the happiness of others as an end and treats every other as an end in itself. It is a community engaged in the harmonious and cooperative pursuit of the good.” (Korsgaard *CKE* 23). A realm of rational agents whose ends form a *system* “when these ends are not only mutually consistent, but also harmonious and reciprocally supportive” (Wood 166). It also might serve as a long-term political and moral goal for humanity, as some interpreters have suggested (Korsgaard *CKE* 153).

2.2 *A deliberative ideal*

Despite the compelling features of the substantive ideal, we have another interpretation of the kingdom of ends as an ideal available to us. Both John Rawls and Thomas E. Hill, Jr. give a deliberative reading of the kingdom of ends which sets the ideal that I (as an individual) may use in order to improve my deliberations.

Rawls uses this way to describe the kingdom of ends in his *Lectures on the History of Moral Philosophy*. Rawls focuses on how the Formula of the Kingdom of Ends (which is closely related to the Formula of Autonomy) is where we “come back to viewing ourselves not as subjects of the moral law, but as legislators, as it were, of the public moral law of a possible realm of ends” (Rawls 204).

¹⁰² Section 5.2.1 will explore imagining the substantive ideal in more detail.

Imagining ourselves as ideal legislators is a very different way in which the kingdom of ends may serve us as an ideal. So perhaps instead of thinking about how *we* ought to act as a collective, we should do the following (Rawls' gloss on the kingdom of ends):

“Always act so that the totality of the maxims from which you act is such that you can regard yourself as enacting through those maxims a unified scheme of public moral precepts the endorsing of which by all reasonable and rational persons is consistent with their humanity and would bring about (under favorable conditions) a realm of ends.” (Rawls 205).

While we could think of the kingdom of ends as capturing a particular moral world, the kingdom of ends is especially effective, we might think, as an ideal decision procedure (Rawls 213). This decision procedure would lead me to check whether or not my maxim would be endorsable by others, and perhaps meet other criteria.

Hill writes that the kingdom of ends serves as a model for “deliberation about moral principles” (Hill *RPJ* 223). We know—by definition—that the kingdom of ends is conceived abstractly as how things would be if everyone did his or her duty (and given Kant's commitment to the importance of a divine actor, if God assisted to make the world such that the ends of the virtuous were satisfied) (Hill *HWMW* 75). However, here are important ways of deliberating about what we should do: we ought to consider ourselves and others as rational, autonomous beings (Hill *RPJ* 223). We should aim to inform ourselves, and to avoid features of our world that would lead us to be less than perfectly rational (Hill *HWMW* n22). While we each have our own ends, we set those aside so we can make legislation together with others (Hill *RPJ* 224). As such, ideal legislators would abstract from their private ends and personal differences (Hill *VRJ* 60, 64ff.)

Hill recommends that we imagine the kinds of things ideal legislators would do as we reflect on how this ideal could guide us. Ideal legislators would (i) seek dialogue with other reasonable moral agents, (ii) confront their problems, and when they cannot resolve their

disputes, act according to the judgments they can best justify to other moral deliberators, (iii) take into account dignity and private reasons (abstractly considered), (iv) review ‘bills’ that are candidates for laws, and (v) pass certain of these bills (Hill *RPJ* 224, 45). Hill writes: “The ideal of moral agents as jointly legislating moral laws, I suggested, urges us to curb our moral self-complacency by consulting others, listening to divergent views, and submitting our own convictions to criticism.” (Hill *RPJ* 45). We aim to deliberate well, and looking to how we would act as such ideal deliberators may serve as guidance for us. Perhaps, too, we should try to create social conditions that empower each agent to have a voice and to hear others regarding moral matters.

Again, while the most obvious way of reading the ideal of the kingdom of ends is as for a moral community, Kant’s ethical writings also could be taken to support a reading of the kingdom of ends as a *deliberative* ideal. We know that the *Groundwork* (4:434, 438) instructs us to consider ourselves as legislators in a kingdom of ends. Perhaps we can take that idea home: look to the ways in which legislation goes best, and think of our ordinary moral deliberation in a similar manner. We would think of the characteristic beliefs, desires, and mental life of a certain kind of person. Doing so would allow us to avoid hastiness, failure to accommodate the needs and ends of others, consider what others need in order to exercise their agency, and get past any kind of complacency.

3. The Uses of an Ideal

We have seen two interpretations of the kingdom of ends as an ideal available to us: as a *substantive ideal* and as a *deliberative ideal*. Both interpretations are grounded in ways in which

Kant describes and discusses the kingdom of ends. At this point, we could continue to ask the exegetical question: which interpretation does the text best support? I suggest, however, that we take on a different approach to the task. Instead, I will ask what kind of purposes ideals serve for us, in line with a broadly Kantian view of morality. If we can identify those purposes, we might be able to consider why we might want to read the kingdom of ends as one or the other kind of ideal.¹⁰³

One reason why we might think that ideals are important and interesting is due to a major concern in Kantian ethics. This concern is about our empirical inclinations—our needs, our fears, and so on. These features of our psychology have a propensity to pull us *away* from what duty commands us to do and *toward* things that we think will make us happy (*G* 4:405). We might wish to be free of these inclinations—as Kant briefly suggests would be best for us (*G* 4:429). However, it is more reasonable to conclude that we cannot be rid of them, and so should decide what to do in the face of our inclinations.

One thing we ought to do is make ourselves *ready to resist* any inclinations that tempt us away from what morality requires (Hill and Zweig 32). However, there might be other options available to us. Reason is the centerpiece of what we ought to do in Kantian ethics and motivates us to act rightly. I assume, for my purposes, that we've settled what reason requires of us, what our duties are, and so on. However, I think that it is important for Kantians to reckon with how human agents are also moved to act by other parts of our psychologies—including our emotions and our concern for consequences. While the moral law's motivational force is placed on us as rational agents, the remainder of our motives could be aligned or disaligned with the direction to which the moral law directs us. It might be worth considering whether or not we can make it

¹⁰³ In other words, I will be considering substantive and deliberative ideals as subsets of the class of imaginative ideals.

such that our inclinations are, as much as possible, aligned with what the moral law requires of us.

I propose that Kantian ethical ideals can do this work to harmonize the rest of our moral psychology with what reason requires.¹⁰⁴ They provide a kind of complementary motivation that speaks to these other parts of ourselves. Ideals (as objects of the imagination) can certainly motivate us—they can inspire us; they encourage us. However, it seems that only ideals *of particular kinds* can move us in ways that are aligned to what the moral law requires of us. Some ideals—like Gauguin’s ideal of artistry—in fact tempt us *away* from what the moral law requires. Yet the ideals that Kant identifies, including the ideal of the kingdom of ends, seem to lead us in a direction that corresponds to our duties. We are attracted to the possible consequence that we could live in a world with such a gracious community, while being able to pursue our projects, all compatible with living in the way that the moral law requires of us. On a Kantian view, ideals do not *motivate* us to act morally. However, they can do the work of making the rest of our inclinations lesser temptations to do other than what reason requires.

Chapter 1 has detailed many of the ways that imaginative ideals can motivate us to act. As such we might be able to intuitively understand why such immersion could overcome temptations to do other than what we ought to. For example, we might consider our typical problem of trying to act in ways that benefit ourselves by making exceptions for ourselves (*G* 4:424). Imagining a moral community of rational agents might require us to attend to the

¹⁰⁴ As described in the Introduction, Kant was criticized for neglecting the role of the emotions in ethics by a set of philosophers ranging from Bernard Williams to a number of feminists. Allen Wood writes that Kant was similarly criticized for neglecting the role of the irrational and of feeling in religion (Wood 201, referring to Lösment’s *Zur Religionsphilosophie Kants* and Otto’s *The Idea of the Holy*). Wood thinks that those criticisms are misled: Kant surely recognized the role played by “the irrational” and “inner feeling” in ecclesiastical faith (Wood 201-2). However, Kant’s aim was to “make a *rational* assessment” of those phenomena (Wood 202).

concerns and decisions made by moral agents beyond ourselves. We respond to imaginings in ways that affect our emotional responses and our real-world judgments.¹⁰⁵

If we are to use ideals as an ethical tool for ourselves, then it does not seem to make our well-fitting emotions a matter of luck or coincidence, but part of what it is to reconcile ourselves to what the moral law requires of us. As such, this alignment of our inclinations is something that we can choose for ourselves, by purposefully choosing to imaginatively engage with ideals (or other methods) to make sure that our inclinations do not pull us away from what reason requires of us.

I acknowledge that I depart from Kant's ethical writings insofar as I will speak of the kingdom of ends as a kind of *imaginative* ideal (in the sense described in Chapter 1), and how it might be able to serve us as such as an ideal. Nonetheless, ideals seem to be a very fruitful part of ethical theory. Kantians should avail themselves of this tool if there is a way that is compatible with the overall thrust of Kantian ethical theory. What I propose is one manner to think through these ideas.

Moreover, it seems like my interpretation does not fall woefully away from Kant's view. Kant says that "[Moral inquiry]¹⁰⁶ serves to produce in us a lively interest in the moral law by means of the splendid ideal of a universal kingdom of ends in themselves (rational beings) to which we can belong as members only if we are scrupulous to conduct ourselves in accordance

¹⁰⁵ See Green and Donahue, "Simulated Worlds: Transportation into Narrative" for a survey of studies done in this area.

¹⁰⁶ The replaced text is "It", but it is not definitively obvious that "moral inquiry" is the referent. I have used 'moral inquiry', as that is the overall topic of the paragraph and the referent for the previous "it" in the paragraph. The only other option—which makes less sense, I believe—is that the "Idea of a pure world of the understanding, as a whole of all intelligences to which we ourselves belong as rational beings" is what produces in us a lively interest. This is a *possible* reading, which perhaps means that Kant is speaking about the relationship between the *Idea* of the kingdom of ends and the *ideal* of the kingdom of ends. Nonetheless, I think that 'moral inquiry' is the better reading.

with maxims of freedom” (*G* 4:462-3). Why might an ideal of the kingdom of ends, a community of virtuous agents, make the moral law more accessible to us? Imagining a kingdom of ends hooks up to “intuition and feeling” (Hill and Zweig 86, ref to *G* 4:436-7). Its appeal is due to how it would represent (abstractly) the world that would be actual if everyone followed the moral law (Hill and Zweig 86).

We know we lack control over what comes about as a result of our dutiful action, partially because consequences cannot be perfectly predicted, and partially because the success of many of our dutiful actions depend on the cooperation of others (or more minimally, a lack of conflict with others). And we are interested in the consequences of our actions, as Kant well recognized (*R* 6:5). Imagining an ideal of a certain kind might be able to ‘fill in’ the story about what the consequences of our actions could be, if others were to also act rightly.

With this taken as our direction, I propose that we can re-think the work of examining the role of the kingdom of ends as a substantive ideal or as a deliberative ideal. I will in more detail explain the ways in which we imagine, and how we imagine both kinds of ideals. While doing so, I will ask how the substantive ideal or deliberative ideal should serve us as a method to align our emotions or concern with consequences with what reason requires of us.

4. The Imagination and Ideals

We can recall from Chapter 1 that imaginings are quite like perceptions or beliefs or desires. They can mimic perceptions, beliefs, and desires (depending upon the case) and sometimes substitute for them (Currie and Ravenscroft 11). Ideals can often represent a complex network of perceptions and beliefs and desires, and our desires *for* them represent deep-seated

desires we have not merely in the context of the imagining, but desires for things that we believe are of incredible value.

We can *imaginatively immerse* in an ideal. In the case of a kingdom of ends as a substantive ideal, we might imagine what it would be like for such a community to come to pass: what strategies they might use to resolve conflicts, or how they justify themselves to each other. Immersion—as Susanna Schellenberg tells us—can be beneficial to learning, allowing us to see perspectives that we might not have otherwise and to practice new strategies (Schellenberg 508).¹⁰⁷ The characteristics of the imagination might be able to illuminate how the kingdom of ends should serve as an tool for us.

4.1 Mirroring and contagion

The imagination has two characteristics that are especially relevant to this discussion. The first is mirroring, in which imaginings accurately reflect real features of the world. The second is quarantining and contagion, the spectrum by which we recognize that the imagining is a mere imagining, or alternatively, ends up influencing our behavior beyond the scope of the imaginative episode. Let us discuss both in more detail.

The first feature, ‘mirroring’, refers to the ways in which imaginings can accurately reflect real features of the world. As Tamar Gendler explains it: when we pretend X is Y, X is taken to have the effects and features *within* the pretense that Y is in reality believed to have (Gendler 142-3). When I imagine a conversation that I’m going to have with a friend, the conversation I imagine would take into account my and my friend’s tendencies in conversation, values, willingness to listen to each other, and so on.

¹⁰⁷ Chapter 1, Section 3.1 described imaginative immersion in more detail.

Imaginative ideals, too, can mirror realistic physical and psychological consequences to the degree that they aptly take on characteristics from the ordinary world. We can ‘see’ ourselves in the kind of ideals that we work with. When we imagine a virtuous person, for example, we imagine a person who has the capacity to feel pain and suffering, who has the temptation to not follow the moral law, or who might have to work in order to understand their own motives. Despite these temptations, they have a good will that is also strong. We would not be able to ‘see’ ourselves as the virtuous person if we were to imagine a person who is not faced with the same kinds of needs and temptations as we are.¹⁰⁸

Likewise, if we imagine the kingdom of ends (on either interpretation), we will be mirroring realistic features of our world. For the community, we would imagine the members of the kingdom of ends to be like us in many ways—they would communicate using speech (rather than the more sci-fi option of having a group mind), they would have a number of private ends, they would be susceptible to temptations, and so on. These are not fully abstract ‘rational agents’ but agents who we can imagine to be relevantly like ourselves. For a legislating agent, we would imagine the same, except the focus would be what it is like to be an individual who is facing a deliberation. This person would need to consult with others, would need to set aside their self-interest (otherwise they might be tempted), and would know that their motivations can sometimes be opaque even to themselves. Much of the content would be the same between the two kinds of ‘mirroring’, but the features of what it is like to engage in a deliberation are much different than what it is like to objectively consider a community.

The second feature of imagining that is worth discussing is that of ‘contagion’. In ordinary cases of imagining, we recognize that a particular pretense episode *is* a pretense

¹⁰⁸ See *R* 6:63-4 for Kant’s comments along these lines.

episode, and anything that we imagine is a mere imagining (Gendler 7). This kind of ‘quarantining’ reminds us that we should not take an imaginative episode to deeply influence what we think is true about the world—just because I could imagine what it would like to be betrayed by my friend does not mean that I *have* been betrayed by her. Nonetheless, despite the way in which we ordinarily quarantine our imaginings, when we are vividly immersed in an imaginative episode, emotions and beliefs that are responses to what we are imagining might ‘escape’ quarantine. So while we might recognize that the imagining we engage with is not reality-reflective, we can’t help ourselves. We engage with it in as if the content *were* reality-reflective (Gendler 242). This can provide a kind of *contagion*.

While contagion in many cases may provide us with trouble—imagining a monster in our closet might have kept our childhood selves up at night, for example—it can also be a benefit. This is especially true with regard to ideals. ‘Contagious’ ideals can be an incredible kind of benefit. In the last chapter, for example, I offered an account of why imagining a virtuous person could have a ‘contagious’ effect on our ordinary lives. So we might want to further discuss what the role of quarantining and contagion is for the kingdom of ends in both its possible interpretations: can imaginative engagement with the ethical community or the ideal legislator have an impact on our actions that we might welcome?

4.2 Testing the interpretations

How do the two interpretations of the kingdom of ends do as imaginative ideals? To what degree can we see ourselves in the ideal? And, given the ways in which we see ourselves mirrored in the ideal, to what degree can it move us to act?

4.2.1 The substantive ideal

In the first case, we would mirror features of the communities we are actually familiar with. This might involve imagining how we converse and share ideas with others, acknowledge the values and priorities of others, and engage in cooperative action of various kinds. We might try to imaginatively immerse in ways in which we have seen community members work with each other well and mutually support each other.

However, there are limits on the ways in which we can mirror real features of the community. There are two limits in particular that are worth mentioning. First, communities are complex. We know, even in small groups of people, that it is difficult to understand and appreciate what others value and would choose to do accordingly. When you extend that complexity to the concept of “all rational agents”, our imagination hits a limit. We might be able to imagine such a community including the features described in the *Groundwork*, or features about human needs (as described by Barbara Herman in section 2.1). However, we might be deeply limited in how we could ‘illustrate’ the ideal community imaginatively. We’re not able to keep in mind what it might be like, except in a very abstract manner, what kinds of *persons* are involved in the community—what they care about or how they might make plans based upon what they care about. In short, the complexity of an ‘ethical community’ means that what we imagine is not particularly vivid.

Second, there may be many possible realizations of what this community would be like. These possible realizations that we could conceive are highly likely to involve mutually inconsistent institutions or priorities.¹⁰⁹ So even to whatever limited degree where we could

¹⁰⁹ Certain ‘ideal communities’ might contain particular institutions that work well within a larger institutional structure. Counterfactually, these ideal communities could have been set up in such a way that they had very different institutions. To take a small-scale example, consider how families divide household labor for egalitarian purposes. One arrangement might be to have each person take half of each chore (switching off days to do the dishes and

imagine these possibilities, even abstractly, there might be no matter of fact as to what ‘the ethical community’ would look like. That would grossly limit the way in which we could imaginatively immerse in this ideal.

The substantive ideal may still be valuable, despite its limits as an imaginative ideal. Given the multiplicity of options available, we might be better able to rule *out* certain kinds of ways in which a community would act using this ideal, rather than imaginatively immerse in a particular conception of a community. We would rule out, for example, communities in which agents only consider a subset of persons and their values in our deliberations about what to do, rather than to consider the community as a whole. We would rule out malicious behaviors and activities, including projects (or formulations of projects) that would undermine others. So, for example, competition might be permissible, but not policies of ‘winning no matter the cost’ (Wood 169-70).

Moreover, imagining the ethical community hooks on to the powerful feelings that we have in wanting to be with others. When we imagine the substantive ideal, doing so might hook on to feelings that we have for others: our love for others, our desire to help others. The ideal of an ethical community could channel those inclinations in a specific way—towards beneficence, generosity, or sympathy. These inclinations could be ‘contagious’—present when we imagine the ethical community, and which also come to mind as we are acting in our ordinary lives. Nonetheless, given our difficulties in imaginatively immersing in this ideal, the patchiness of our imagining would mean that it would have limitations to *how* it could direct our inclinations.

laundry). A different arrangement might be to have each person to take the chores that they like or don’t mind doing (one partner always does the dishes because they hate laundry, and the other partner dislikes laundry less). These might be, once the systems are fully evaluated, part of equally ‘ideal’ systems for the household. However, those practices could not be put into place for the household simultaneously.

Second, imagining that such a world is *possible* may secure a kind of hope for us. When we imaginatively immerse in the ideal of a harmonious moral community, we might be inspired by our feelings of pleasure in thinking of a community of persons who are like us, but who have managed to live well together. Considering those attractive features of the ideal can help to avert our despair. Negative feedback—evidence that others do not act rightly, worries that others may harm us—might lead us to give up on the desire to establish an ethical community. However, mentally contrasting an imagined future to our current reality has been shown to be a strategy to enable us to respond to such negative feedback (Oettingen and Kappes 395). This mental contrasting functions leads us to be energized to put in efforts to bring about the desired outcome, and commit to plans that would allow us to attain the desired future (Oettingen and Kappes 396-7). This is an important function for the substantive ideal.

4.2.2 The deliberative ideal

The deliberative ideal is a much easier ideal for us to imagine. It would mirror various aspects of what our own deliberation is like—how we consult others in order to see perspectives unlike our own, or how we do well when we are willing to subject our judgments to the evaluations of others. We might go so far as to imagine rational agents who are susceptible to cognitive biases or the impact of emotions, and so ideal deliberators of this kind must take due care to avoid their influence and establish safeguards.

The ideal of an excellent deliberator is easy to ‘see’ ourselves in, and we can easily imagine ourselves behaving in ways that the deliberative ideal entails. We’re more familiar with what makes our deliberations successful than what it is to make a community morally perfect. Moreover, bringing ourselves to be more like ideal deliberators seems to be a matter that is more

in our own control, rather than something that requires the coordination and cooperation of others.

Deliberating well may also show how we have reason to want to be with others not only because of some empirical facts about our sociality (e.g., it makes us happy to be around others), but also because of truths about how we exhibit rationality. We do not come into the world as fully rational agents, but instead develop into mature rational agents. Part of what it takes to be a successful rational agent is to be able to deliberate with others, and be able to justify our deliberations to others.

Andrews Reath's "Legislating for a Kingdom of Ends: The Social Dimension of Autonomy" makes a point along these lines. Autonomy, for us, comes with features relating to our lives among rational agents, including that we have "the capacity to formulate and act from reasons and principles that can justify one's actions to other rational agents" (Reath 224). In other words, we are capable of giving others reasons to accept or endorse our choices (Reath 225). The deliberative procedures that we use are ones "by which one can create reasons that bind other agents" (Reath 225). Our procedures create permissions, rights and duties, or confer value, and which others must acknowledge (Reath 225). Understood in this way, autonomy is a power "exercised in relation to others, made possible by their responses and requiring their participation" (Reath 226). Our exercise of our autonomy is limited by the possibility of others' sharing one's conclusions (Reath 227). As such, the deliberative ideal may not be only able to align our intuitions with what reason requires, but to also be a matter of being an autonomous person among other rational agents.

The deliberative ideal is contagious, too, but in a different way than the substantive ideal. It does not catch on to the robust feelings that we have about wanting to live with others.

Nonetheless, it does hook onto our calmer desires to be able to deliberate well. Moreover, we know how powerful, for example, to imagine a version of *ourselves* (Hershfield). The impact of imagining ourselves as a better deliberator has an incredible kind of effect on real-world motivation and performance as a result.

In short, the deliberative ideal is more imaginable than the substantive ideal, and highly likely to impact our practices of deliberation if we are to imaginatively engage with it. As such, the deliberative ideal might be able to serve as a kind of method to give us a sense of what it would be like to be a person living within such the ‘ethical community’ that we hope for. The substantive ideal of the ethical community can spur us to hope. However, it does not give us precise instruction for *how* to live with others. For that, perhaps we must look to the deliberative ideal. We may then, want to think of the deliberative ideal as its counterpart: a strategy for acting well, with others.

What does this tell us that we should imaginatively engage with, then? Both of our ideals. The substantive ideal of the moral community can elicit hope of a certain kind. The deliberative ideal is ‘contagious’ in a way that the substantive ideal is not. Having both available to us can give us a tool for aligning our inclinations toward a hoped-for community (which would mean meeting our duties would bring such a world about), and a method that is within each of our powers to exercise.

Conclusion

This paper has done a few things. It introduced Kant's discussion of the kingdom of ends as an ideal. Then, it considered how interpreters have read the ideal of the kingdom in two ways: as a substantive ideal and as a deliberative ideal. This paper then looked at the ways in which those two readings of the ideal could serve different roles, and how we need both to serve very different roles for the kingdom of ends to have its full effect as an imaginative ideal.

The conclusion that we should be led to is that the two ideals serve our purposes very differently, and we should engage with both, for their different matters. Our use of the substantive ideal of the kingdom of ends is a balm to our worries that *perhaps* we will not be able to live well with others, despite our hopes. Nonetheless, the deliberative ideal serves us best as guide to action in our ordinary deliberations and practicing our autonomy well with others.

WORKS CITED

- Adams, Robert Merrihew. "Introduction" to *Religion within the Boundaries of Mere Reason*, by Immanuel Kant and edited by Allen Wood and George di Giovanni. New York: Cambridge University Press, 1998.
- Currie, Gregory and Ian Ravenscroft. *Recreative Minds: imagination in philosophy and psychology*. New York: Oxford University Press, 2002.
- Dean, Richard. "Humanity as an Idea, as an Ideal, and as an End in Itself" in *Kantian Review*, Vol. 18, No. 2 (2013), pp. 171-195.
- Flikschuh, Katrin. "Kant's Kingdom of Ends: Metaphysical, Not Political" in *Kant's Groundwork of the Metaphysics of Morals: A Critical Guide*, edited by Jens Timmermann. New York: Cambridge University Press, 2009.
- Herman, Barbara. "A Cosmopolitan Kingdom of Ends" in *Moral Literacy*. Cambridge, MA: Harvard University Press, 2007.
- Hill, Jr., Thomas E. *Dignity and Practical Reason*. Ithaca, NY: Cornell University Press, 1992.
- Hill, Jr., Thomas E. *Human Welfare and Moral Worth*. Oxford: Clarendon Press, 2002.
- Hill, Jr., Thomas E. *Respect, Pluralism, and Justice*. New York: Oxford University Press, 2000.
- Hill, Jr., Thomas E. *Virtue, Rules, and Justice*. New York: Oxford University Press, 2012.
- Hill Jr., Thomas E. and Arnulf Zweig, "Editors' Introduction" to *Groundwork for the Metaphysics of Morals* by Immanuel Kant, edited by Thomas E. Hill, Jr. and Arnulf Zweig. New York: Oxford University Press, 2002.
- Gendler, Tamar. *Intuition, Imagination, and Philosophical Methodology*. New York: Oxford University Press, 2010.
- Green, Melanie C. and John K. Donahue. "Simulated Worlds: Transportation into Narratives", Chapter 16 of *The Handbook of Imagination and Mental Simulation*, edited by Keith D. Markman, William M. P. Klein, and Julie A. Suhr. New York: Psychology Press, 2009.
- Hershfield, Hal E. et. al. "Increasing Saving Behavior Through Age-Progressed Renderings of the Future Self" in *Journal of Marketing Research*, Vol. XLVIII (Special Issue 2011), pp. S23-S37.
- Kant, Immanuel. *Critique of Pure Reason [C1]*, translated by Norman Kemp Smith. London: Palgrave Macmillan, 2007.
- Kant, Immanuel. *Critique of Practical Reason [C2]*, translated by Lewis White Beck. New York: Macmillan Publishing Company, 1956.

- Kant, Immanuel. *Critique of the Power of Judgment* [C3], edited by Paul Guyer, translated by Paul Guyer and Eric Matthews. New York: Cambridge University Press, 2000.
- Kant, Immanuel. *Groundwork for the Metaphysics of Morals* [G], translated by Arnulf Zweig, edited by Thomas E. Hill Jr. and Arnulf Zweig. New York: Oxford University Press, 2002.
- Kant, Immanuel. *Lectures on Ethics* [LE], edited by Peter Heath and J.B. Schneewind, translated by Peter Heath. New York: Cambridge University Press, 2001.
- Kant, Immanuel. *Metaphysics of Morals* [MM], translated and edited by Mary Gregor. New York: Cambridge University Press, 1996.
- Kant, Immanuel. *Religion within the Boundaries of Mere Reason* [R], edited by Allen Wood and George di Giovanni. New York: Cambridge University Press, 1998.
- Korsgaard, Christine M. *Creating the Kingdom of Ends*. New York: Cambridge University Press, 1996.
- Korsgaard, Christine M. “Introduction” to *Groundwork of the Metaphysics of Morals* by Immanuel Kant and edited by Mary Gregor and Jens Timmerman. New York: Cambridge University Press, 2012.
- Moran, Kate. *Community and Progress in Kant’s Moral Philosophy*. Washington, DC: The Catholic University of America Press, 2012.
- Oettingen, Gabriele and Andreas Kappes, “Mental Contrasting of the Future and Reality to Master Negative Feedback”, Chapter 26 of *The Handbook of Imagination and Mental Simulation*, edited by Keith D. Markman, William M. P. Klein, and Julie A. Suhr. New York: Psychology Press, 2009.
- O’Neill, Onora. *Constructions of Reason*. New York: Cambridge University Press, 1989.
- Rawls, John. *Lectures on the History of Moral Philosophy*. Cambridge, MA: Harvard University Press, 2000.
- Reath, Andrews. “Legislating for the Kingdom of Ends: The Social Dimension of Autonomy” In *Reclaiming the History of Ethics: Essays for John Rawls*, edited by Andrews Reath, Barbara Herman and Christine M. Korsgaard. New York: Cambridge University Press, 1997. EBook. Cambridge Core. Accessed March 8, 2017.
- Schellenberg, Susanna. “Belief and Desire in Imagination and Immersion”, in *The Journal of Philosophy*, Vol. 110, No. 9 (2013), pp. 497-517.
- Wood, Allen W. *Kant’s Moral Religion*. Ithaca, NY: Cornell University Press, 1970.

CHAPTER 5: FRIENDSHIP AS A BRIDGE TO THE HIGHEST GOOD

The last chapter gave an argument that we could treat the kingdom of ends as a substantive ideal for a moral community—one that would give us hope that we could live well with others. The chapter also gave a solution for how we could feel that *achieving* such a community was within our power: to use the legislator from the kingdom of ends as a deliberative ideal. If each of us were to make our own deliberations as much like an ideal deliberator as possible, perhaps that would be one way to bring our world closer to the kingdom of ends. Nonetheless, we may still have found this to be only, at best, a partial solution. The deliberative ideal is addressed to each person as an individual. We may be concerned that our method for bringing about a community has very little to do with community-making.

These concerns are even more pressing against a backdrop of other criticisms about how Kant addresses our sociality. There has been a strong tradition of dismissing Kantian ethics because it doesn't make the right kind of room for intimate attachments. One line of this familiar criticism is that Kantian ethics has an overly-strong requirement of impartiality. This requirement of impartiality would mean that either those using the theory are inappropriately blind to those we love when we are impartially deciding what to do, *or* risk having 'one thought too many' when we decide that this is an occasion in which we can prioritize assisting our loved one (Williams "PCM" 2).¹¹⁰ A related challenge has come from feminist critics of Kant, who say

¹¹⁰ Another version of this criticism can be illustrated by Michael Stocker's famous case of a person who goes to the hospital to visit a friend, and in response to the friend's gratitude, explains that he is "only doing his duty" (Stocker 462).

that the kinds of things that we associate with intimate relationships—choices guided by emotion, for instance—are left out of the Kantian ethical picture. Carol Gilligan’s *In a Different Voice* drew out the ways in which the moral reasoning of men and boys differed from the moral reasoning of women and girls. Some readers of the Gilligan study took it to be a call against Kantian ethics: Annette Baier wrote, for example, that “the emphasis in Kantian theories on rational control of emotions, rather than on cultivating desirable forms of emotion, is challenged by Gilligan” (Baier 30).¹¹¹

Kant’s discussion of what the ideal of friendship is and how it should be acted upon in our lives may provide the Kantian with some kinds of responses to these criticisms. This chapter will do a few things: it will introduce what Kant says¹¹² about the ideal¹¹³ of friendship. Then it will examine Kate Moran’s take on the duty of friendship. Moran takes Kant’s discussion of friendship to allow for a satisfying response to those critics who think Kantians cannot appropriately account for social relationships. Moreover, she thinks that Kantian agents *must* engage in friendships of certain kinds (Moran 170). The final two sections of the chapter will

¹¹¹ This setup of the two criticisms is from Moran (169). Velleman’s “Love as a Moral Emotion” (339n3) has a similar list of interlocutors.

¹¹² There are three passages in the corpus where Kant extensively discusses friendship. The first is a section at the end of the Doctrine of Virtue where Kant talks about the ideal of friendship. The second source is the Collins lecture on ethics, where Kant is described as giving a description of the kinds of friendship that he takes there to be: the friendship of need, the friendship of taste, and moral friendship (*LE* Collins 27:423-4). Third, he discusses friendship in the Vigilantus lectures (*LE* Vigilantus 27:675-83).

The *Lectures on Ethics* are dated to before the publication of the *Metaphysics of Morals* (1797) (Sullivan vii). The earlier Collins lectures probably reflect what Kant taught in his ethics courses from 1775-80 or so (Schneewind xvi). They come close to some of Kant’s mature views without presenting Kant’s characteristic views (like the Categorical Imperative) (Schneewind xvi). The later Vigilantus lectures reflect a course that Kant gave in 1793-4, so a period much closer to the publication of the *Metaphysics of Morals* (Schneewind xviii).

¹¹³ The *Metaphysics of Morals* refers to friendship as an ideal, but we must look to the Collins lectures for his characteristic distinction, again, between the ‘idea’ of friendship and the ‘ideal’ of friendship. The idea is a measure or “maximum” of what friendship could be. The ‘ideal’ is insofar as that idea can serve as a pattern (*LE* Collins 27:423).

partially agree with the Moran view, but suggest that her account has missed important work done by the ideal of friendship.

In particular, I will argue that Moran's view needs two extensions. She argues that moral friendship is a duty because it allows us to bring about our moral self-perfection (Moran 199-200). I argue in Section 3 that it does more than that: moral friendship *is both* in the service of our imperfect duty of self-perfection, *and also* is in the service of our imperfect duty of helping others. If my interpretation is correct, then friendship is an ideal that, in microcosm, shows what a social system where we both bring about our own perfection *and* the happiness of others would be like. This view will be explored in Section 4. As such, moral friendship—which can, importantly, actually be achieved ‘here and there’—does an incredible amount to secure our hopes for what is possible as a result of our moral activity. Moral friendships help to alleviate our doubts that we cannot, with others, bring about something relevantly like the highest good.

1. Kant on Friendship

Let us begin by exploring the claims that Kant makes about friendship. In the *Metaphysics of Morals*, Kant describes the ideal of friendship as “the union of two persons through equal mutual love and respect” (*MM* 6:469). The first part—a union of two persons—is obvious to us. The second part—that of equal love and respect—requires a bit more unpacking. On the face of it, those claims seem innocuous: after all, most of us want both love and respect in our intimate personal relationships. However, the way in which Kant describes (in particular) the features of respect in a friendship leave some readers cold.

The section in question reads as follows:

“how can he be sure that if the *love* of one [friend] is stronger, he may not, just because of this, forfeit something of the other’s *respect*, so that it will be difficult for both to bring love and respect subjectively into that equal balance required for friendship? – For love can be regarded as attraction and respect as repulsion, and if the principle of love bids friends to draw closer, the principle of respect requires them to stay at a proper distance from each other.” (MM 6:470).

Readers generally have no problems with love being as an attraction to another person, and love as a central part of friendship. Respect as Kant describes it is less intuitive: why would we want to be *drawn away* from the other person?

Part of the reason why most readers have this reaction is because when ordinary readers consider what they want in a loving relationship, they already have a conception where appropriate limitations are *built in*. I may love my partner and love my best friend, but I certainly do not want them to be within 10 feet of me all the time nor would want them to read every email or text I send. Love—in Kant’s sense—is without those built-in boundaries. It is an unchecked interest in the other person: “to feel in possession of each other in a way that approaches fusion into one person” (MM 6:471). Love is good, but it is not bounded in Kant’s sense—it is solely the attraction we have to the other person.¹¹⁴

We need to balance out this attraction that we have to the other person with something else. This something, for Kant, is respect. And even though it is unattractively described as ‘repulsion’ in the passage, we might think of it in more intuitive terms as a kind of limiting condition on how intimate we choose to be with a given person at a given time.¹¹⁵

¹¹⁴ My reading is similar to that found in Moran 173-4 and Denis 5-6.

¹¹⁵ The Collins lectures speak of friendship very differently than the later Vigilantus lectures and the *Metaphysics of Morals*—as a “maximum of mutual love”. In these early lectures, friendship is framed around love for others and happiness (LE Collins 27:423). We might find it interesting that there is no discussion of respect, yet. In contrast, the later Vigilantus lectures do contain this emphasis on both love and respect (LE Vigilantus 27:682), and is much more closely aligned with the description from the *Metaphysics of Morals*.

The next claim that Kant makes about friendship is that friendship makes us *deserving* of happiness (but does not *secure* happiness for us) (*MM* 6:469). What might this claim mean? We already know what it takes to be worthy of happiness: to be virtuous (*G* 4:393). That makes our question the following: how does friendship make us virtuous? Virtue is where an agent has a will that is not only good but is also strong. She is able to resist motivations (e.g., self-interest) that threaten to override her motivation to do as the moral law demands (*MM* 6:380, 405; *LE* Mrongrovius 29:627). Engaging in friendships then, must help us develop a *good* will, a *strong* good will, or both.

One reason that we might not do as we know that we ought to is because of our natural mistrust of others. One kind of reason that we mistrust others is because we fear their judgment (*MM* 6:471-2). We are, of course, imperfect people—revealing those things about ourselves that are less admirable is likely to lead an observer to think less of us. If that is the case, we fear that they will respect us less. A second kind of reason that we mistrust others is that the more they know about us, the more they could harm us.¹¹⁶ Betrayal, for instance, is more potent when it is done from a position of information—knowing what the betrayed person most values. As a result of these kinds of concerns, we might be tempted to isolate ourselves from other people (Baron 368, 371).

Friendship is an “aid in overcoming the constraint that we harbor, from mistrust, towards those we associate with, and in opening up to them without reserve” (*LE* Collins 27:428). With a trusted friend, we can communicate our thoughts and sentiments without fear that the person will share our secrets (or otherwise harm us), or that the disclosure will undermine our mutual respect

¹¹⁶ Moran will add a third reason to this list: our tendency to compare ourselves to others and to seek superiority (Moran 177). This will be explored slightly later in this section.

(Baron 375-6). While we may remain mistrustful of others in general, with regards to our *friends* we can be secure in our trust. As Denis writes, “Kant praises friendship because he understands it as a relationship that embodies love and respect for others, preserves self-respect, and fosters self-development” (Denis 8). If so, having such friendships might be an important part of becoming virtuous.

Friendship allows us a special opportunity—it allows us “to participate in a relationship of mutual aid and enjoyment” (Moran 177). Nonetheless, we should not assume that Kant thinks that moral friendships are easy (or even possible!) to achieve. He mentions two difficulties in particular: (a) that it is difficult to ensure that each person has equal and reciprocal feelings of love and respect for each other (*MM* 6:469), and (b) it is equally difficult to ensure that love and respect themselves are in appropriate balance.

Kant thinks that friendships are best when the friends have *equal* reciprocal love and respect for each other. However, achieving that equality is difficult, due to a third tendency of ours that tends to pull us away from what the moral law requires. Kant points out that we are the kind of creatures who are inclined to compare ourselves to others and to seek superiority to others (Moran 176-7, referring to *R* 6:27). This means that we will have a tendency to seek superiority (not equality) even over our friends (Moran 177). In becoming close to a person, we become particularly vulnerable to vices of comparison and competition because we know more about them (Moran 177). Moreover, because of the difficulty in knowing the truth about others’ feelings (and sometimes the truth about our own feelings) we might have an epistemic barrier to ensuring that friendships are equal.

Second, ensuring that love and respect are themselves balanced also seems difficult, if not impossible (*MM* 6:469-70). Kant seems especially worried about the following scenario. One part of friendship is to help our friend by pointing out their faults. In a perfect friendship, the friend will not take offense because he understands that this is a duty of friendship (Baron 376)—but given our susceptibility to wanting to compare ourselves with others, it is likely that many times things will not go so well. Moreover, our continued tendency towards competition with others may us feel *pride* in ourselves, and a kind of contempt for the person who has revealed themselves to us. Thus reflecting on their faults may lead us to lose respect for our friends (Baron 373).

So, according to Kant's views about friendship, we might think that it is difficult to achieve or even unattainable. There is some dispute about whether we should take it to be unattainable or not, because Kant's remarks are somewhat equivocal. He says that, like many of our other ideals, "it is readily seen that friendship is [...] unattainable in practice, although striving for friendship [...] is a duty set by reason" (*MM* 6:469). Yet slightly afterward, he says that friendship "actually exists here and there in its perfection" (*MM* 6:472). Some interpreters have resolved this apparent contradiction by suggesting we pull apart two senses of friendship: that when he speaks of friendship as *unrealizable*, he is talking about an ideal of "perfect friendship". When he speaks of it being *rare but realizable* he speaks of "moral friendship" (Baron 394, Korsgaard 199).¹¹⁷ We will return to this question in Section 4.3.

¹¹⁷ Baron writes that the text is ambiguous and does not decide the matter, but thinks that it is likely that 'moral friendship' and 'perfect friendship' are not identical (Baron 374). Moran and I treat "perfect friendship" and "moral friendship" as the same concept, but I will return to this alternative interpretation later on in the paper.

1.1 Kinds of Friendships

There are two other tasks to take on in an introduction to Kant on friendship: first, to discuss what Kant means by contrasting “moral friendships” with what he calls “pragmatic friendships”. Second, we should also discuss Kant’s tripartite division of friendships from the Collins *Lectures on Ethics*.

Kant wants us to make a distinction between moral friendships and friendships based on feeling (*MM* 6:471). In the *Metaphysics of Morals*, Kant describes this friendship of feeling as a ‘pragmatic friendship’: in which we take on the ends of others from love (*MM* 6:472).¹¹⁸ We might think that a pragmatic friendship is the “union aimed at mutual advantage” that Kant mentioned earlier in the section (*MM* 6:470).

We may be able to understand a bit more about pragmatic friendships if we look to Kant’s earlier description of friendship from the Collins *Lectures on Ethics*. In those lectures, Kant is described as distinguishing between three types of friendships: the friendship of need, friendship of taste, and the friendship of disposition.¹¹⁹

A *friendship of need* is a special relationship out of self-interest (LE Collins 27:425). These sound very much like ‘unions aiming at mutual advantage’—where “participants may entrust each other with a reciprocal concern in regards to their needs in life” (LE Collins 27:424-5). We could perhaps imagine Hobbesian actors in a state of nature forming these kinds of

¹¹⁸ Kant also mentions ‘the friend of humanity’ distinct from his discussion of moral or perfect friendship (*MM* 6:472). This friend takes an ‘affective’ interest in the well-being of others (*MM* 6:472). Given Kant’s description of this friend as having an ‘affective’ interest, we might think that this is an analogue of the “pragmatic friendship”. However, Kant’s description of the friend of humanity seems to be much more a counterpart of the moral friendship. The ‘friend of humanity’ is concerned with the equality between persons, and seems to be a relation of both mutual love and respect (*MM* 6:472-3).

¹¹⁹ As Lara Denis notes, these three categories of friendship are taken from Aristotle (*Nicomachean Ethics*, Book 8, Chs 3-6) (Denis 3).

friendships. Or consider a friendship formed by two people who don't have friends in a new city, and so eat lunch with each other so they won't be alone. We should think of these friendships as based on reciprocal mutual aid.

A friendship of taste is a little more sophisticated than the friendship of need: it is where friends take pleasure in mutual association (LE Collins 27:426). Moran writes that what makes this type of relationship so valuable to us is that each party can learn from the other—such friendships broaden our horizons (Moran 182). I might appreciate a friend because of their excellent music taste, another friend because of their witty conversation, another friend because they know a lot about the cinema. On Kant's view, the reason I find these friends so valuable is because they have some skill that I do not possess. It seems likely that we should think of this kind of friendship as another kind of "pragmatic friendship": this time where each friend supplies a want of *expertise*.

A friendship of disposition, in contrast to the first two kinds of friendship, seems to be the same thing as what Kant described as "moral friendship". Like in the *Metaphysics of Morals*, Kant describes this friendship of disposition as rare and founded on the friends being concerned with the same principles of morality (LE Collins 27:429). Likewise, such a friendship is motivated by how we want to be social with one another, but are fearful of what may come (LE Collins 27:426-7). Such a friendship is the way in which we can securely allow "the complete confidence of two persons in revealing their secret judgments and feelings to each other, as far as such disclosures are consistent with mutual respect" (MM 6:471).

Putting an exact count to the kinds of friendships that Kant categorizes is difficult. We would need to decide (i) whether we think there is a distinction between 'perfect friendship' and 'moral friendship' in the *Metaphysics of Morals* (which I and Moran do not use, but other Kant

scholars have), (ii) whether we want to treat the three notions of friendship in the Collins lectures as canonical, and (iii) if so, whether the ‘pragmatic friendship’ lines up with a friendship or taste, or need, or both (as I have assumed).¹²⁰ I set these concerns aside in order to focus on Moran’s interpretation of the ideal of friendship.

2. Moran on the Duty of Friendship

In this section, I will introduce Kate Moran’s account of the duty of friendship. Many Kantians have been motivated to respond to the critics (described in the Introduction) who think that Kant’s ethical system has no room or inappropriate room for intimate attachments (Moran 168). While Moran agreed with other Kantians that friendship does have a role in the system, she disagreed with their strategy. Those Kantians argued that friendship is one way to engage with our imperfect duties (Moran 169-70). Yet on that account, Moran does not see how a Kantian agent has any kind of *obligation* to seek out relationships like friendship (Moran 170). If friendships are one way of discharging our imperfect duties, well and good—but it seems like friendship is perhaps optional. I could discharge my imperfect duties in some other way and not do anything wrong. As Moran says, “[a] person could, in other words, be a perfectly moral person and never participate in a relationship like friendship.” (Moran 170).

¹²⁰ Denis counts four notions of friendship: first, friendships of taste, need, and disposition (the same as moral friendship) (see *LE* Collins 27:425-6). These three kinds of friendship are realizable (Denis 3). The fourth kind of friendship, perfect friendship, is not realizable (Denis 3). It incorporates elements of the other kinds, particularly the characteristics of moral friendship (Denis 3). Moran seems to take there to be three kinds of friendship: the ‘friendship of disposition’ from the Collins lectures is the same as “moral friendship” (and she does not distinguish between ‘moral’ and ‘perfect’ friendship) (Moran 178ff).

In contrast to those accounts, Moran thinks that moral friendship plays a unique role in our lives as a necessary condition to develop a moral character (Moran 171). As such, we have a duty to participate (or to try to participate) in moral friendships (Moran 170).¹²¹ Moral friendship, Moran explains, “is a kind of friendship that exists between two parties who are less than perfectly moral, yet who strive to become moral” (Moran 184). Thanks to their trust and comfort with each other, friends can share their feelings and thoughts with each other (Moran 184-5). This trust and comfort will allow, Moran thinks, for us to help to morally perfect ourselves.

An important duty in Kantian ethics is our duty to know ourselves (Moran 186, referring to *MM* 6:441). Unfortunately, we also have a tendency toward self-deception (Moran 188). We can do some things as individuals to try to mitigate this tendency. We can engage, for instance, in thought experiments of “imagining how one’s actions might affect others, or how one’s actions might come into a conflict with the ends of others” (Moran 193, referring to *C3* 5:293-5).

However, Kant also thinks that we will not be able to secure the kind of objectivity needed for moral self-examination on our own. We have to share our thoughts with others to secure ourselves against errors (Moran 186, 88, referring to *LE* Collins 27:427, *LE* Vigilantus 27:616, 27:683, and *MM* 6:469). A trusted friend can provide a needed external perspective to help us examine our actions and provide an extra buttress against self-

¹²¹ Kant explicitly calls friendship a ‘duty’ (*MM* 6:469). He does not explain exactly why, but we might wish to guess that it depends on how moral friendships can make us virtuous (even if he does not clearly explain how he thinks that they do make us virtuous). It is unclear from the passage alone as to what this means to guide our action: whether we have a duty to engage with our existing friends in a certain way, whether a person should make friends if they have none, or so on.

deception (Moran 192).¹²² Friends who have known us for a longer period are, moreover, especially attuned to the idiosyncratic mistakes that each of us tend to make about ourselves (Moran 191). We know that we have certain moral bad habits—they come with the territory of being human. We tend towards arrogance in some instances, self-contempt in others (Moran 189). Both are ways in which we can deceive ourselves about our moral worth (MM 6:441).

If Moran is right—and I believe she is—we can develop both a good and strong will not only because friendships help us avert our natural mistrust of others, but also because friendships are a necessary condition for our self-examination to be successful.

Thus, we can understand what Kant mean when he said that we have a duty to participate (or try to participate) in moral friendships (Moran 170). This duty to pursue character-improving friendships stems from our duty to pursue the highest good. Again: the highest good is a state of universal virtue, and correspondingly universal happiness (C2 5:129). We have a duty to perfect our own virtue (part of working toward the highest good) and so we have a duty to take part in activities that improve our moral character, including friendship with some person (Moran 178). As Kant points out in the *Groundwork*, to will an end means also to will the means to that end” (Moran 200). Moral friendship is a necessary means to the end of moral self-perfection. Thus, we must engage in moral friendships with others.

Moral friendships help us to perfect ourselves morally. In contrast, Moran reads the other two kinds of friendships (of need and of taste) as ways in which Kantian agents can practice their imperfect duties to others (Moran 171). The friendship of need is, on Moran’s reading, an opportunity to practice our imperfect duty of beneficence (Moran 179). I promote my friend’s

¹²² Nonetheless, Kant thinks that each of us must be careful about how we counsel our friends if we see them tending toward one or the other vice: we don’t want to point out the moral failings of our friend in a way that compromises their self-respect. Tact is required (LE Vigilantus 27:685, MM 6:470).

happiness because I am engaged in mutually aiding her. The friendship of taste, where we take pleasure in the company of others who we can learn from, is a way to fulfill our duty to perfect our talents. Moran argues that this is because we are poised to learn from other people who are dissimilar to us (Moran 183).¹²³ I might become a wittier conversationalist by speaking with a clever friend, or develop a better taste in music by listening along with my music-loving friend. Thus, the other two kinds of friendship are opportunities to practice our remaining imperfect duties. The moral friendship is solely directed toward our moral self-perfection (Moran 199).

I admire Moran's move to respond to critics who think that the Kantian moral agent is poorly socialized.¹²⁴ I also think that she is right insofar as friendships are an essential means to perfecting ourselves, as one part of our duty of bringing about the highest good. Her arguments about the friendship of need and the friendship of taste are deeply interesting.

However, there are two things the ideal of friendship serves to do that are missed by Moran's account: *first*, the ideal promotes the duty to bring about the happiness of others *as well as* the perfection of each, meaning that it is a small-scale version of the highest good. The highest good is the good that we, as human agents, necessarily set as an end for ourselves: in which we attain moral perfection, and receive proportionally deserved happiness (*R* 6:4-5, 6-7n). *Second*, Kant finally gives us room to think that the ideal (or something quite like the ideal) is realizable with regard to friendship—meaning that perhaps moral friendships can serve as a

¹²³ Both of these kinds of friendship have their dangers, according to Moran. Our dispositions to seek superiority over others, though, might lead us to destroy our friendships of need. So we must practice this duty without “a sense of pride, resentment, or vanity” (Moran 181). The friendships of taste are also dangerous: where we become jealous of our friends' talents or take too much pride in our own accomplishments (Moran 183).

¹²⁴ Certainly this is not what Kant intended: he writes that “The human being is a being meant for society [...] and in cultivating the social state he feels strongly the need to *reveal* himself to others” (*MM* 6:471). He follows the section on friendship with one in which he says “It is a duty to oneself as well as to others not to *isolate* oneself” (*MM* 6:473).

bridge to the highest good, and help us avoid despair about how such a good may be out of reach. In the remainder of the chapter, I will set out arguments for each of these views.

3. Moral Friendship and the Imperfect Duties

We should agree with Moran's argument: friends are incredibly good at helping each other identify what is true about themselves, and in doing so, can assist their friend's pursuit of moral perfection (*MM* 6:470). Her view also gives us a more robust view of why engaging in friendship makes us 'deserving of happiness'—that is, virtuous. Part of the project of becoming virtuous is inspecting our own motivations for action and avoiding self-deception. A trusted friend can assist me in this project (so far as we are careful to act in ways that do not compromise our respect for each other).

Yet I disagree with Moran's view that we should take moral friendship as *primarily* concerned with self-perfection (Moran 199). I think we have reason to think that moral friendship is not merely aimed at perfecting our characters, but also is in the service of mutually assisting each other—and as such, is directed toward our duty of beneficence.¹²⁵

In order to understand why this is a better reading, let us return to how Kant describes the ideal of friendship. Kant writes that it is:

“an ideal of each participating and sharing sympathetically in the other's well-being through the morally good will that unites them, and even though it does not produce the complete happiness of life, the adoption of this ideal in their disposition toward each other makes them deserving of happiness” (*MM* 6:469).

¹²⁵ I am not alone in this interpretation. Christine Korsgaard writes that friends care for the other's happiness instead of his own (Korsgaard 191).

This description should indicate to us that both of our imperfect duties are in play. Not only do we do what makes us “deserving of happiness”—to develop and act on a good will—we ‘participate and share’ in the other’s well-being. It may be the case that our activities do not, ultimately, produce happiness—after all, Kant considers happiness to be a very fickle pursuit.¹²⁶ Nonetheless, it is reasonable to suppose that in addition to promoting our self-perfection, friendships also promote others’ happiness.

This seems in line with what we ordinarily take friends to do for each other: to promote each other’s well-being in a variety of ways. Moral friendship is more than sharing our inner judgments and feelings but is a matter of reciprocal concern. The earlier *Lectures on Ethics* support this claim. In the Collins lectures, Kant describes friendship as very much tied to happiness. Particularly, friendship is concerned with reciprocity in cultivating happiness (*LE* Collins 27:422-3). Likewise, the Vigilantius lectures indicate that well-being springs from moral friendships (*LE* Vigilantius 27:682). If this is the case, then friendships are relationships in which we can simultaneously, and properly, pursue both of our imperfect duties at once.

Friends are both in a special kind of position to scrutinize each other’s motives and also to promote each other’s permissible ends. Our friends should not be the exclusive beneficiaries of our work to assist others, but we might think that we can assist them in special kinds of ways, given our intimate knowledge of their projects. A friend who knows our secret anxieties about an upcoming talk might offer to listen to a practice run, and one who has been entrusted with the fact that we have Bieber fever might take us to a concert for our birthday. We can assist our friend in promoting their ends in a way that is according to their own conception.

¹²⁶ As Kant writes in the *Groundwork*, “the concept of perfect happiness is such a vague concept that although everyone wants it, they can never say definitively and self-consistently what it really is that they wish and will” (*G* 4:418).

We also can help our friend refine their conception of happiness. We have many desires that would not lead to our overall well-being. We also have desires that could lead to our happiness, but are mutually inconsistent or jointly unrealizable. Kant thinks that we must bring our inclinations and desires into some kind of tolerable system that is subordinated to, and consistent with, what the moral law requires (*C2 5:73*). If our friend is placing too much emphasis on the importance of being thought well of (at the expense of what they authentically care about), we can point that out—which might assist them, in turn, to refine their conception of what they truly value. Or if they are behaving in ways that are selfish, we too can point that out to them.¹²⁷

If this is a promising reading of the moral friendship, then we have reason to believe that moral friendships are special kinds of opportunities to both perfect ourselves morally and non-morally as well as to practice beneficence. Correspondingly, we should also re-think Moran's interpretation that the other two kinds of friendships (e.g., friendships of need and friendships of taste) are ways in which we can practice the remaining imperfect duties (cultivating our non-moral talents and practicing beneficence, respectively).

The distinction between actions in accordance with duty and actions *from* duty may be helpful here. If a person would like to cheat someone but decides not to in case they are found out, we know that they are acting in line with duty ('not cheating') but not from duty (their

¹²⁷ I have focused on the ways in which beneficence might be part of moral friendship, because of the textual support given (and also, ultimately, my purpose in trying to show that moral friendship can provide a kind of bridge to the highest good). We might also be able to imagine how moral friendships could be able to support our development of non-moral talents, however, instead of solely our moral self-perfection.

Moran, for example, thinks of the friendship of taste as perhaps going beyond an opportunity as an opportunity to perfect our talents, but also "By associating with different types of people, we naturally cultivate our ability to think from other perspectives and to imagine how our actions might be consistent—or inconsistent—with their ends." (Moran 193).

motivations are due to their anxieties about what would happen *if* they cheated). Similarly, a person might aid someone solely because they happen to like how it makes them feel when they help someone. Such actions are ‘in accordance’ with what duty requires of us, but (based upon how we have framed these cases) they are not motivated by the moral law. In contrast, an agent who acts in certain ways because they know it is what the moral law requires (or constrains their behavior purposefully based upon what the moral law requires) would be acting *from* duty. A person might be motivated by both affections for another person as well as from duty. In that case, their maxim for action in line with their affections require would still need to be consistent with what the moral law requires.

I suggest that the friendships of need and taste are ways in which we can engage in practices that are often in accordance with duty (toward beneficence and perfecting ourselves). However, the source of our motivations in these pragmatic friendships are not necessarily tied to duty. They are instead tied to mutual advantage—of love of our friend of some variety. Again, friends would be acting in accordance with duty, but not *from* duty, when they permissibly practice beneficence or cultivate their talents. And since their motivations are not dutiful ones, their non-moral motivations might distort their actions. I might end up prioritizing my friend (a person I love) over other people who might be able to use my assistance. I might begin to *use* my friend who has talents very different than mine as a mere means to perfecting myself. I might try to promote my friend’s happiness, but impose my conception of happiness on her instead of being sensitive to the choices she has made.

This means that while I can work on certain things in friendships of need and friendships of taste—being attentive to what would make my friend happy, or ways in which I have a capacity that needs to be cultivated—the motives are untethered to the moral law unless we

engage in moral friendship. *Moral* friendship is where we unite what we may have learned from our friendships of need and friendships of taste. They are not susceptible to the same kinds of potentially non-moral behavior. And as friendships based on equal and reciprocal love and respect for each other, our beneficent or self-cultivating actions are from duty (and perhaps as well from love for our friend,¹²⁸ or self-respect). My beneficence to my friend is limited by my respect for her. I choose to promote her end because it is *hers*—a moral agent's—who I have particular knowledge of. My friend might be able to provide me with insight into how to develop my non-moral talents. Like in the case of offering me suggestions about my motivations for action, they would give me feedback in a way that fundamentally is in line with their respect for me.

In short, moral friendship is based upon *respect* for the friend as a rational agent in a way that the friendship of need and the friendship of taste are not. Friendships of need and friendships of taste are based on conditional kinds of reciprocity—that I have been beneficent to my friend, and my friend has been beneficent to me; or that my friend has talents that I do not, and vice versa. If one of us ‘drops the ball’, then the friendship unravels. In the case of a moral friendship, we unconditionally respect the person as a moral agent. That respect for our friend, combined with the love we have for our friend (which makes the friendship particular) means that the imperfect duties both can be practiced in the context of a moral friendship—but in a very different way than in the other two kinds of friendship.

¹²⁸ I here follow Kant's discussion about moral motivation, insofar as we can be motivated to act both because of moral law and empirical incentives, so long as those empirical incentives are ‘subordinated’ to the moral law (*R* 6:36). So my loving desire that my friend be happy no matter what, must be subordinated to the direction of the moral law about the permissibility of her desires, and her pursuit of her own perfection.

3.1 An Objection

We should carefully consider why Moran insisted that moral friendship was directed towards morally perfecting ourselves and not aimed at beneficence. Surely, we might think, promoting our friend's well-being (at least to some degree) is a natural part of being their friend. So why would these friends not be concerned with bringing about the well-being of their friend?

One reason that Moran may be convinced that moral friendships are not also directed toward beneficence is because Kant seems to warn us of dangers of helping our friend in the context of a moral friendship. There is an odd passage in which Kant is concerned about how assisting a friend in securing their advantage might result in the loss of respect for the friend (*MM* 6:470-1). I want my friend to be better off than she is, so I help her. But if I do this, he suggests, I may lose my respect for her, because I see that she is unable to help herself (*MM* 6:470). As a result, Kant says that in a moral friendship, your friend is someone who you can always rely on, but you never *do* rely on—if you have a burden, you even should conceal that burden from your friend (*MM* 6:471). So there is textual reason to think that perhaps Kant thought that moral friendship was much more about moral self-perfection than it was about beneficence.

Nonetheless, I do not think we should not take Kant's worry about this possibility as reason to think that beneficence is not properly part of moral friendship. We should instead take his concern as a call for *due care*. Marcia Baron's interpretation of that odd passage of Kant is that he is referring to talking about *financial* burdens, not emotional or other kinds of burdens (Baron 377). As he says in the *Vigilantius* lectures, asking for financial assistance from a friend might lead to the weakening of the friendship (*LE Vigilantus* 27:684). However, there are many non-financial opportunities to promote my friend's well-being. I can come to his art show, or

spend time enjoying a walk with him (something that promotes both of our well-being!) These kinds of activities do not seem to lead me to lessen my respect for my friend. Moreover, there is an important difference between soliciting our friend for help in a particular direction, and the friend choosing freely to assist us in a certain way that they know will promote our happiness.

That is very different from refraining from ever promoting each other's well-being. Kant only says that the assistance that we can expect from the friend should not be the "end and determining ground of friendship" (*MM* 6:470). However, the reason why I might be friends with a person might be based upon my love and respect for her—and it only happens that I choose to fulfil my duty of beneficence (among other ways) to assist her on some occasions, when I judge that it is consistent with what the moral law requires.

We want our friend to be happy, because happiness is a conditional good predicated on their being virtuous. And part of being engaged in a moral friendship is to assist each other in their pursuit of becoming more virtuous. Nonetheless, we should not take moral friendships to involve sitting on our hands with regard to our friend's happiness. We should just be careful that we do not fall away from the primary concern of maintaining both love *and* respect for our friend in equal shares.

4. The Ideal of Moral Friendship as a Bridge to the Highest Good

So far, I have accepted Moran's reason to interpret moral friendship as a means to self-perfection. I have also added a claim that moral friendship is also a way of fulfilling our duties of beneficence with regard to a person that we have a special kind of intimate knowledge of. Doing so comes with various kinds of dangers that we must be careful to avoid. Nonetheless, the ideal

of friendship is a relationship between persons that is aimed at the virtue of the friends, as well as (but secondarily) their happiness. This description of the ideal of friendship—as involving both virtue and happiness—might start to remind us very closely of the highest good. As Stephen Engstrom explains, the concept of the highest good is “the idea of a state or condition in which happiness is proportioned to virtue” (Engstrom 747).

One kind of frustration we might have with the ideals that Kant sets up in his moral philosophy—including the ideal of the highest good—is that he both says that they are *unrealizable* and that we must work to attain them. For some of our ideals—like the ideal of virtue and the ideal of the highest good—not only must we work to attain them, we must believe that we *can* attain them (*R* 6:66-7, *C2* 5:113). As we have seen in prior chapters, that can also lead to unexpected commitments—like needing to posit the existence of God and our immortal souls (*C2* 121-122/126-7).¹²⁹

However, we have seen that moral friendships are realizable. If so, this means that an ideal closely related to the highest good is pursuable even if we know that most persons in this world may not be virtuous, and that conditions in this world are often apt to not promote happiness (much less happiness that is perfectly proportioned to virtue). In what follows, I will explore how we might be able to think of moral friendships as a kind of way in which we can bridge to the highest good, in a way that again does not rely on metaphysical assumptions that we may not wish to follow.

¹²⁹ With the length of time afforded by an immortal soul, we could continue to perfect ourselves; with God’s assistance, we would be granted the amount of happiness that we deserve. If we are not inclined to make those metaphysical commitments, we may despair that the highest good could ever come about.

4.1 The Relationship between the Kingdom of Ends and the Highest Good

The highest good and the kingdom of ends are closely related concepts. Some interpreters have even thought that they referring to the *same* concept. Allen Wood, for example, speaks of the ethical community of the kingdom of ends as the same as the highest good (Wood 1970 188-9). Likewise, Moran says that the highest good went under revisions in Kant's various writings, but in its 'final and most complete form', we are "presented with a moral system that has as its goal a kind of moral community whose members are governed by the moral law" (Moran 25).

Nonetheless, we should not think of the highest good and the kingdom of ends as identical. To follow a distinction from Barbara Herman, we might want to articulate a difference between a kingdom of good wills and the highest good (which involves good wills as well as happiness) (Herman 64). The reason that this difference is important is because of the highest good requires happiness to be both *proportional* to virtue, and also that that proportionality is not coincidental, but is necessary (Engstrom 750).

A community of morally good human agents could not ensure on their own that the happiness of each person is necessarily and rightly in proportion to their virtue. At best we might be able to promote each other's happiness to the best of our ability. We can do various kinds of work in social cooperation to try to achieve something very like the highest good (see Engstrom's remarks on this topic, pp. 778-9). Nonetheless, we shouldn't think that the kingdom of ends is necessarily the same thing as the highest good comes about (unless we have built a divine legislator into our account of the kingdom of ends).

For these reasons, we might think of the kingdom of ends as the humanly-achievable portion of the highest good. We may not be able to ensure necessary proportionality of happiness to each person's virtue, but we can work to promote each other's happiness. (For anything more,

we would need those metaphysical postulates of Kant's moral religion.) This means that the kingdom of ends has quite an important role for us—it sets an ideal that is what we can achieve cooperatively together.

4.2 The Imaginative Ideal of Friendship and the Kingdom of Ends

With that distinction between the kingdom of ends and the highest good in mind, we might ask how friendship might connect us to the humanly-achievable portion of the highest good. Our frustration in the previous chapter might be due to only having a partial solution as to how we could bring about the kingdom of ends. Surely each of us could aim to act like the ideal deliberator, and in so doing act in ways that (were we all to do so) would bring about the kingdom of ends. Nonetheless, we might find it frustrating that we were not engaging with others who do the same. Moran's discussion of the ideal of friendship indicates a second worry: that even if we were to each seek to be more like the ideal deliberator, that would only do so much—we would lack an external perspective if we were to not converse with a trusted friend to whom we could reveal the full scope of our reasoning.

Perhaps we now have a solution to our initial frustration: a vividly-imaginable ideal that also engages in the relationship between persons. That is what I will explore in this section: the ideal of friendship as the complement to the ideal deliberator, where both are needed in order to secure our hope that we could bring about a community like a kingdom of ends.

The ideal of moral friendship is a two-person version of the highest good, wherein each person is made virtuous and correspondingly happy. We can imagine what it would be like for each of us to be in a friendship *like that*—it would have all the features of the best friendships. As such, it can have all the benefits of other imaginative ideals—being 'contagious'

in our ordinary lives. I can vividly imagine what it would be like if I and my friend were like the moral friends, and that imagining may spill over into how I engage with my friend.

Second, we can imagine how those moral friendships bridge to the highest good as a version of the kingdom of ends. We can imagine many persons engaging in moral friendships, or moral friendships between multiple persons. This would be like adding links on the chain so it wasn't just a community of two, but a community of many. Such a community would be a kingdom of ends on earth; a place in which the highest good could come about (were conditions to be such that happiness was proportioned to virtue). Perhaps this can certify why we ought not despair that the highest good could not come to pass. Such a concept brings ideals to earth; we can see that they are—perhaps—achievable, and how they would be achievable.

Other Kantians have argued that friendship is in the direction of the kingdom of ends. Allen Wood argues that friendship is the clearest model in human life for the kingdom of ends (Wood 1999 279). Christine Korsgaard argues that “Friendships are human moral achievements that are lovely in themselves and testify to the virtue of those who sustain them. To become friends is to create a neighborhood where the Kingdom of Ends is real” (Korsgaard 194). If we are right, then the ways in which we engage in friendships—a trust of others, and a reciprocity that's not based upon contingent features of our world, but respect for the other person as a moral agent—then can be extended to other members of my community.¹³⁰

In short, our task is to become a legislator-friend. We should try to engage in friendships that are as much like the moral friendship as possible—where we balance our love for our friend

¹³⁰ In the *Metaphysics of Morals*, Kant discussed a “friend of humanity” who gives “thought and consideration for the equality” of persons (*MM* 6:473). Marcia Baron, in her explanation of that passage of the *Metaphysics of Morals*, says that Kant takes friendship to have two proper forms: (1) the union of two persons through equal mutual love and respect, and (2) being a “friend of human beings as such” (Baron 366; *MM* 6:469, 472). Perhaps in Kant's discussion of the second type of friendship, we can see how the highest good can result from the friendship between two people turning each into more of a “friend of human being as such”.

with respect, and where (so far as we can) we try to have an equal and reciprocal relationship with each other. Doing so will allow the friends to be able to advise and provide an external perspective for their friend as a deliberator, and promote each person's virtue (and their corresponding happiness). We can (and should) deliberate with non-friends by seeking their views and sharing our judgments, but we know that we are safest when we do so with our friends. As such, we should take our deliberations with our friends to form a model for our work engaging with other members of our community—and should seek to engage with more persons in ways that are relevantly like moral friendships.

4.3 Is the Ideal of Friendship Realizable?

One reason that we might be especially interested the ideal of friendship is because it feels closer to being 'realizable' than the other ideals he mentions in his moral writings. Remember, there are two available interpretations of Kant's description in the *Metaphysics of Morals*: that the ideal of moral friendship 'actually exists, here and there' (MM 6:472). Or, he thinks that there is an ideal of 'perfect friendship' that is unattainable, but there is another kind of friendship (a "moral friendship") that actually exists here and there. I argue that on either interpretation, moral friendship helps to buttress us against our potential despair that the highest good is unrealizable.

4.3.1 Interpretation 1: The Ideal is Achievable

There are a few reasons to interpret the ideal of moral friendship as itself achievable. The opening description of the "perfect friendship" and the later statement of the "moral friendship" are very similar to each other—in both cases, he asks us to consider a kind of friendship "in its perfection". Nor does Kant ever contrast 'perfect' and 'moral' friendship. However, when Kant

says that “This (merely moral friendship) is not just an ideal but (like black swans) actually exists here and there in its perfection” (*MM* 6:472), it might seem inconsistent with his earlier claim that friendship is ‘unattainable in practice’ (*MM* 6:469). Yet perhaps there is some plausible explanation for this. Kant might be purposefully overstating the claim of ‘unattainability’ as a way of making us aware of the many difficulties in securing such a friendship (*MM* 6:470).

If he should be read as saying that the *ideal* is achievable, then this is a radical change. The only other ideal that he has granted as having been attained is the ideal of moral perfection of the Son of God—a clear reference to Christ or a Christ-figure (*R* 6:61). However, we might think that that ideal is a very different order of ‘attainable’—one that is out of reach of ordinary persons. In contrast, the friendships he seems to be referring to ‘here and there’ are ones that would be between persons who don’t have a special theological standing. So perhaps this *achievable* kind of ideal might certify our hopes that the kingdom of ends could come about.

We might, too, want to reflect on why Kant compared moral friendships to ‘black swans’. Black swans were once used (as in Juvenal’s *Satires*) as metaphor for things that did not exist. However, after Europeans discovered in the 1600s that black swans *did* exist, the meaning shifted to indicate that we may be surprised to discover that something *does* exist, and which now proves the wrongness of our wide-ranging assumptions in the first place. So similarly, we might think that the surprise of finding out the moral friendships *can* exist here and there—despite of all our shortcomings—should lead us to think that our assumptions that we could have some kind of social life (at least a two-person one) that is in correspondence with what duty requires.

4.3.2 Interpretation 2: The Ideal is Unattainable, but Moral Friendship is Attainable

Our other available interpretation of the passage on friendship in the *Metaphysics of Morals* is that there are two kinds of friendship at play. Moral friendship is as good as it gets, and though rare is not non-existent (Baron 374). In moral friendship, the form of reciprocity one has is in frank conversation and sharing of sentiments (Korsgaard 199). Perfect friendship, in contrast, is characterized by feelings of equal mutual love and respect (Korsgaard 199)—and is the ideal.¹³¹

This interpretation of the passage on friendship does make a lot of interpretive sense. It explains why Kant calls “perfect friendship” an ideal, and “moral friendship” ‘no ideal’ (*MS* 6:472).¹³² It also makes “perfect friendship” consistent with all the other ideals, which are described as unattainable within human lifetimes (*MM* 6:406, 409; *R* 6:122, 135). Finally, it resolves the apparent contradiction in Kant’s text by reading him as saying that moral friendship “actually exists here and there in its perfection” (*MM* 6:472) while perfect friendship is unattainable in practice (*MM* 6:469) (Baron 374).

Why might Kant want to distinguish between the ideal form of friendship and exemplary friendships that are real? One kind of reason is simply that he is very careful about how he wanted to talk about ‘ideals’. More precisely, Kant wanted to avoid using particular details in framing an ‘ideal’. He was worried that if we talked about ideals in stories or anecdotes, we might be led astray by how we have conceived of the ideal (*LE* Mrongrovius 29:605). So perhaps

¹³¹ Baron takes this explanation to also explain why Kant referred to moral friendship as “merely” moral friendship (Baron 374, referring to *MM* 6:472). I find this claim to be less persuasive, as earlier in the *Metaphysics of Morals* Kant had an entire section dedicated to duties to ourselves “merely as a moral being” (*MM* 6:428). I do not think it is wise, then, to take the word ‘mere’ as a signifier of a lack of importance or standing in the theory.

¹³² The original German text says that moral friendship is “kein Ideal”. The translator of the Cambridge Edition, Mary Gregor, translated in favor of “moral” and “perfect” friendship being identical by rendering the passage as saying that moral friendship is “not just an ideal” (*MM* 6:472).

we should take his insistence on the difference between “moral friendship” and “perfect friendship” as trying to avoid tying particular persons and the features of their friendships to the ideal of friendship.

If this is the better reading of the passage, Kant’s discussion still allows moral friendship to play the counterpart role to the deliberative ideal, and which together certify our hope that the kingdom of ends could come about. Moral friendship is still described as persons revealing their private feelings and judgments to each other (so far as is compatible with mutual respect) (*MM* 6:471). If so, these friends are still able to provide each other with the kind of external perspective that the other needs and in other ways can assist their friend to become more virtuous. They also can act beneficently toward each other. Moral friendships still can capture our attention as ways in which people have created something like a small-scale versions of the kingdom of ends, even if they still involve imperfect persons and disproportionate happiness.

Thus, I take it to be reasonable to think that “moral friendship”, on either interpretation, is sufficient to help justify our hopes that the kingdom of ends is perhaps achievable for creatures like us. On the second interpretation, we are led to a less dramatic rendering of this possibility, but still we retain reason to hope.

4.4 An Objection

One worry that we might have about taking moral friendship as a bridge to the kingdom of ends is that we can’t be friends with everyone (*LE* Collins 27:430). Friendship is by its very nature a partial relationship. We may exhibit a general good-will to everyone in a moral community, and in doing so would act in ways that are *impartial*. If this is the case, then we might think that there might not be something relevantly like ‘friendship’ in the kingdom of ends. Why, then, would we think that friendship can get us *to* the kingdom of ends?

However, I argue that certain duties are ‘scaffolding duties’ to the kingdom of ends. Some of these involve things like forgiveness (which can only happen after a wrong has taken place), and are a kind of imperfect duty in recognition of the fact that we are all imperfect and need each other’s forbearance on occasion. Likewise, having political institutions that punish wrongdoers are other arrangements that help catch the people who choose not to act morally.

Friendship is one such scaffolding duty. Friendship is a duty for us because we are morally imperfect (Moran 199, *LE* Collins 27:428) and we live among others who are untrustworthy. As such, like a scaffold, we wouldn’t see anything relevantly like it in a community of agents who were virtuous and worked to see each other’s well-being being brought about. Certainly, Kant seems to think this—that we “[start] from the particular and [go] on to the general” (*LE* Collins 27:430). Nonetheless, we can see why partial friendships are important for creatures like us, in social circumstances like ours. Since it is focused on the relationships between persons, it forms a model of how we might be able to trust each other and deliberate with each other, and perhaps give us direction toward a kingdom of ends.¹³³

Conclusion

In this paper, I argued that Moran’s argument for how the friendship is a necessary part to become virtuous needs two supplements: *first*, that we should conceive of the ideal of friendship as also incorporating our duty to bring about each other’s happiness, and *second* to recognize that

¹³³ If a kingdom of ends were to be brought about, perhaps some kinds of partial relationships might still be possible or desirable, but they would look very different from our ordinary partial relationships. We might have intimate knowledge of our friends or loved ones, but not need that intimate knowledge in order to avert our fear of how others may harm us. Nor would we tend towards unjustly privileging only members of our ‘in-group’ in how we conduct our beneficence, as we ordinarily tend to do.

the ideal of friendship is a bridge to the kingdom of ends (which is the humanly-achievable portion of the highest good). Friendship is deeply important for us, as a way to respond to our mistrust of others and develop ways in which we can trust others.

With regard to friendships, we might think that this bridging function of friendships helps respond to worries we might have about *partiality*. Because friendships are partial, they explain the special knowledge we have about our friends and concern we have for them. They are also grounded on mutual respect. Yet unlike other conceptions of friendship, when we engage with a friendship we must ensure that we do not close our hearts towards others (*LE Collins 27:428*). We can love our friends, but not at the expense of concern for others who are not our friends.

WORKS CITED

- Aristotle, *Nicomachean Ethics*, translated by Terence Irwin. Indianapolis, IN: Hackett Publishing, 1999.
- Baier, Annette. *Moral Prejudices*. Cambridge, MA: Harvard University Press, 1994.
- Baron, Marcia. "Friendship, Duties Regarding Specific Conditions of Persons, and the Virtues of Social Intercourse" in *Kant's "Tugendlehre"*, edited by Trampota, Andreas, Oliver Sensen, and Jens Timmerman. Boston: De Gruyter, 2013. ProQuest ebrary. Web. 26 May 2016.
- Denis, Lara. "From Friendship to Marriage: Revising Kant" in *Philosophy and Phenomenological Research*, Vol. 63, No. 1 (July 2001), pp. 1-28.
- Engstrom, Stephen. "The Concept of the Highest Good in Kant's Moral Theory" in *Philosophy and Phenomenological Research*, Vol. 52, No. 4 (December 1992), pp. 747-780.
- Gilligan, Carol. *In a Different Voice*. Cambridge, MA: Harvard University Press, 1982.
- Herman, Barbara. "A Cosmopolitan Kingdom of Ends" in *Moral Literacy*. Cambridge, MA: Harvard University Press, 2007.
- Kant, Immanuel. *Critique of Practical Reason* [C2], translated by Lewis White Beck. New York: Macmillan Publishing Company, 1956.
- Kant, Immanuel. *Critique of the Power of Judgment* [C3], edited by Paul Guyer, translated by Paul Guyer and Eric Matthews. New York: Cambridge University Press, 2000.
- Kant, Immanuel. *Die Metaphysik der Sitten* [MS] in "Das Bonner Kant-Korpus". *Institute für Kommunikationsforschung und Phonetik*, Universität Bonn. Online: korpora.org Accessed March 21, 2017.
- Kant, Immanuel. *Groundwork for the Metaphysics of Morals* [G], translated by Arnulf Zweig, edited by Thomas E. Hill Jr. and Arnulf Zweig. New York: Oxford University Press, 2002.
- Kant, Immanuel. *Lectures on Ethics* [LE], edited by Peter Heath and J. B. Schneewind, translated by Peter Heath. New York: Cambridge University Press, 1997.
- Kant, Immanuel. *Metaphysics of Morals* [MM], translated by Mary Gregor. New York: Cambridge University Press, 1996.
- Kant, Immanuel. *Religion within the Boundaries of Mere Reason* [R], edited by Allen Wood and George di Giovanni. New York: Cambridge University Press, 1998.
- Korsgaard, Christine M. *Creating the Kingdom of Ends*. New York: Cambridge University Press, 1999.

- Moran, Kate. "Friendship and Moral Improvement" in *Community and Progress in Kant's Moral Philosophy*. Washington, DC: The Catholic University of America Press, 2012.
- Schneewind, J. B. "Introduction" to *Lectures on Ethics*, edited by Peter Heath. New York: Cambridge University Press, 1997.
- Stocker, Michael. "The Schizophrenia of Modern Ethical Theories" in *The Journal of Philosophy*, Vol. 73, No. 14 (August 1976), pp. 453-466.
- Sullivan, Roger J. "Introduction" to the *Metaphysics of Morals*, edited by Mary Gregor. New York: Cambridge University Press, 1996.
- Velleman, J. David. "Love as a Moral Emotion" in *Ethics*, Vol. 109, No. 2 (January 1999), pp. 338-374.
- Williams, Bernard. "Persons, character, and morality" in *Moral Luck*. New York: Cambridge University Press, 1981.
- Wood, Allen W. *Kant's Ethical Thought*. New York: Cambridge University Press, 1999.
- Wood, Allen W. *Kant's Moral Religion*. Ithaca, NY: Cornell University Press, 1970.